UNIVERSITY OF ALBERTA

Cost Estimation for the City of Edmonton's Water and Sewer Installation services using an Artificial Neural Network Model.

By

©

DINU PHILIP ALEX

A thesis submitted to the Faculty of Graduate Studies and Research in partial
fulfillment of the requirements for the degree of Master of Science
In
Construction Engineering and Management

DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING
EDMONTON, ALBERTA
Fall 2006

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-22219-5
Our file Notre référence
ISBN: 978-0-494-22219-5

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

ABSTRACT

The construction industry in Canada has been following a dynamic change over recent years. Cost estimation is integral to the success of the construction industry. Over the last six years, the City of Edmonton's Drainage and Maintenance Department has seen a marginal increase of about 12% in the installation of water and sewer services for residential occupancies in Edmonton. The current estimating procedure has showed discrepancies between the estimated and actual cost sustained during the course of the projects to the effect of 60%, and in some cases, the excess of 60%.

This research investigates the factors that affect this variation in costs between estimates and actual costs within the existing process. A detailed analysis of all activities involved in the installation of the water and sewer service installations have been carried out.

The proposed methodology is based on the analysis of all the past data that has been obtained from the City of Edmonton's Drainage Section for the period of 1999 to 2004. The proposed methodology has been incorporated into the module, which integrates the Artificial Neural Network (ANN) and the current estimating system used by the City of Edmonton, 'SmartEST'. This thesis will focus on describing the algorithm used in the ANN and will assess the past data obtained for over 800 jobs (cases) performed over the period of the study.

ACKNOWLEDGEMENTS

"I think that at one point of time everyone appreciates that there are extraordinary men and women and extraordinary moments when history leaps forward on the backs of these individuals............that what can be imagined can be achieved......... that, one must dare to dream......... but in all......... there is no substitute for perseverance, hard work and definitely team work......... because NO ONE gets there alone!!!"

It is with a feeling of profound gratitude and immense regard that I acknowledge and appreciate the valuable and expert guidance of my supervisor, Dr. Mohamed Al- Hussein, whose timely suggestions and ideas have been very instrumental in giving direction to the successful completion of my research project. I would also like to thank him for being more than just a guide, but a mentor throughout my tenure as a Graduate Research Assistant with the Construction Engineering Management group at the University of Alberta.

I am also greatly indebted to Dr. Ahmed Bouferguène from the Campus Saint-Jean who, despite his busy schedule, always managed to find time to suggest valuable inputs during the course of the research work. Special thanks to Siri Fernando and Lino Donata from the City of Edmonton who were always willing to assist in any matters related to the project. I would also like to express my heartfelt gratitude to all the Professors at the University and the officials from the City of Edmonton for all the help and co-operation.

To my parents and my brother, I would like to express a special gratitude without whose encouragement, confidence, support and prayers, nothing would have been possible.

There are many friends and colleagues involved who have been a driving force in making this research what it is. I would like to take this opportunity to thank each and everyone one of them who were involved directly or indirectly in the humble completion of my research.

Thank you!!

TABLE OF CONTENTS

Chapter 1: Introduction 1
- 1.1 Research Motivation 1
- 1.2 Research Objectives 3
- 1.3 Report Organization 4

Chapter 2: Literature Review 5
- 2.1 Introduction 5
- 2.2 Background of Neural Networks 5
- 2.3 Types of Artificial Neural Network Architecture 12
- 2.4 Learning and Training Process of Artificial Neural Networks 14
- 2.5 ANN Application in the Construction Industry 17
- 2.6 ANN Application for Cost Estimation 25

Chapter 3: Current Estimating Practice: The Case of the City of Edmonton 31
- 3.1 Introduction 31
- 3.2 Estimation Software currently used 32
- 3.3 Current Estimation Procedure 32
- 3.4 Results of Current Estimation Procedure 33

Chapter 4: Proposed Methodology 36
- 4.1 Introduction 36
- 4.2 Proposed Research Methodology 36
- 4.3 Data Extraction 38
- 4.4 Data Analysis 41
- 4.5 Modeling Objectives 52

Chapter 5: Development of the ANN Model 54
- 5.1 Why use Artificial Neural Networks? 54
- 5.2 Defining the Supervised Network 55
- 5.3 Data Extraction 56
- 5.5 Network Architecture Selection 57
- 5.6 Output Utilization 62

Chapter 6: Observations 63
- 6.1 Observations 63

6.2 Selected Pattern Analysis 69
6.3 Summary and Comments 73

Chapter 7: Summary and Conclusions 74

7.1 Conclusions 74
7.2 Research Contributions 74
7.3 Recommendations for the Artificial Neural Network Model 76
7.4 Recommendation for Future Research 76

References 78

Appendix A 82

Appendix B 85

Appendix C 104

LIST OF TABLES

Table 1.1: Total Project Cost during the Period of the Study……………………………..3

Table 4.1: Raw Data for a given Job offer……………………………………………….39

Table 4.2: Data formatted during an earlier study……………………………………….40

Table 4.3: Data formatted for easy Comprehension……………………………………..41

Table 4.4: Forecasted Temperature for 2005 based on Yu,H., et al. 2005…..……46

Table 4.5: Neural Network Input Factors and Data types…………………………….52

Table 6.1: Network Results for 3:1:1 Pattern…………………………………………….66

Table 6.2: Network Results for 3:2:1 Pattern…………………………………………….68

Table 6.3: Evaluation of Architectures of different patterns, 3:1:1 & 3:2:1..........70

LIST OF FIGURES

Figure 1.1: Cost Variation for Service Installations for July 1999 to Dec 2004..... 2
Figure 1.2: Variation between Evenly and Unevenly Distributed Cost Estimates . 2
Figure 2.1: Information Transfer via a Synapse 7
Figure 2.2: Basic Artificial Neuron (The McCullogh-Pitts Model) 8
Figure 2.3: Typical Network Architecture 9
Figure 2.4 Sigmoid Transfer Function 10
Figure 2.5: A Simple Feed-Forward Network 13
Figure 2.6: A Feedback Network (Jordan – Elman type network) 14
Figure 2.7: Supervised Learning System 15
Figure 2.8 Back Propagation in Neural Networks 16
Figure 3.1: Service Installations Cost Percentage from July 1999 to Dec 2004... 34
Figure 3.2: Estimated Costs for Service Installations- 2004 35
Figure 3.3: Number of Service Installations - 2004 35
Figure 4.1: Research Methodology Flow Chart 38
Figure 4.2: Variation in Estimated and Actual Costs – February 2004 43
Figure 4.3: Variation in Estimated and Actual Costs – September 2004** 43
Figure 4.4: Cost Variation – December 2004 44
Figure 4.5: Cost Variation – December 2003 44
Figure 4.6: Cost variation – Dec 2002 (1st half) 44
Figure 4.7: Cost Variation – Dec 2001(2nd half) 44
Figure 4.8: Cost Variation – December 2001 44
Figure 4.9: Cost Variation – December 2000 44
Figure 4.10: Cost Variation – Dec 1999 (1st Half) 44
Figure 4.11: Cost Variation – Dec 1999 (2nd Half) 44
Figure 4.12: Monthly Weather from 1953 – 2003 45
Figure 4.13: Cost Variation based on Estimators 47
Figure 4.14: Cost Estimate by Estimator 1 – 2004 48
Figure 4.15: Cost Estimate by Estimator 2 – 2004 48
Figure 4.16: Geographical Distribution: City of Edmonton 49
Figure 4.17: North East Sector 50

Figure 4.18: North West Sector 50
Figure 4.19: South East Sector 51
Figure 4.20: South West Sector 51
Figure 4.21: Modeling Objectives Process Flow Diagram 53
Figure 5.1: Pattern Distribution 57
Figure 5.2: Ward Net – Three Hidden Slabs (layer): Network Architecture 58
Figure 5.3 Learning Process of the Network 60
Figure 5.4: Graphical Representation of Training Sets and Intervals Elapsed 61
Figure 5.5: Contributing Factors and Their Weights 61
Figure 6.1: Five Layer Standard Connection 63
Figure 6.2: Three Hidden Slabs (Ward Nets) 63
Figure 6.3: Output Layer Dampened Feedback 64
Figure 6.4: Hidden Layer Dampened Feedback 64
Figure 6.5: Two Hidden Slabs with Different Activation Functions 65
Figure 6.6: Five Layer, Jump Connection 65
Figure 6.7: Five Layer Standard Connection 70
Figure 6.8: Output Layer Dampened Feedback 70
Figure 6.9: Hidden Layer Dampened Feedback 71
Figure 6.10: Two Hidden Slabs, Different Activation functions 71
Figure 6.11: Three Hidden Slabs, Ward Nets 71
Figure 6.12: Five Layers, Jump Connection 71
Figure 6.13: Prediction Data Plotted Against the Actual and Estimated Values .. 72
Figure 6.14: Predicted Estimates of Previous Cost Variations over 60% 73

Chapter 1: Introduction

1.1 Research Motivation

The City of Edmonton's Drainage Design and Construction Department has many responsibilities, including the installation of water and sewer services from City owned properties to the user's facilities. The City of Edmonton has been utilizing the SmartEST estimation software, which was designed by the University of Alberta's Construction Engineering and Management Group, for the estimation of costs related to installation of the water and sewer services. SmartEST has been integrated with the City of Edmonton's 'Integrated Data Management System' (IDMS) and the 'Systems Applications Products in Data Processing' (SAP). The last six years, from July 1999 to December 2004, has shown a marginal increase of 12% in the installation of such services. Based on the data that has been obtained from the IDMS and SAP, the current estimating technique has showed discrepancies between the estimated and actual costs to the effect of 60%, and in some cases, the excess of 60%. Figure 1.1 demonstrates the percentage variation between the estimated and actual costs pertaining to the water and sewer service installations during the period of the study (July 1999 to December 2004). Such an inaccuracy in the estimate is unacceptable by the City organization.

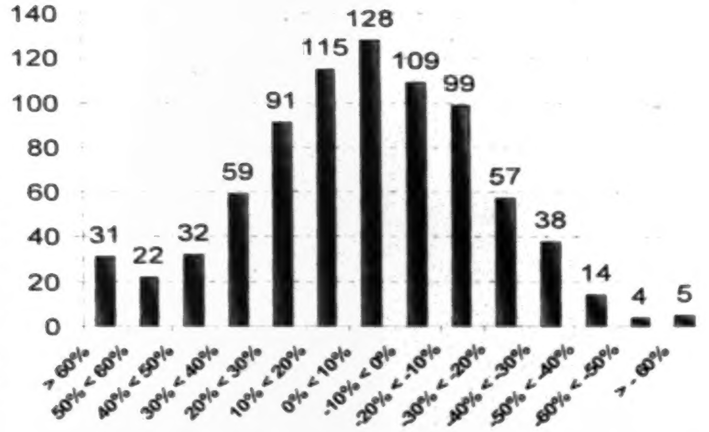


Figure 1.1: Cost Variation for Service Installations for July 1999 to Dec 2004

Only 29% of the projects fall within the ±10% level of accuracy that the City organization targets as illustrated in Figure 1.2. A monthly breakdown of the service installations during each year over the period of the study is shown in Appendix A.

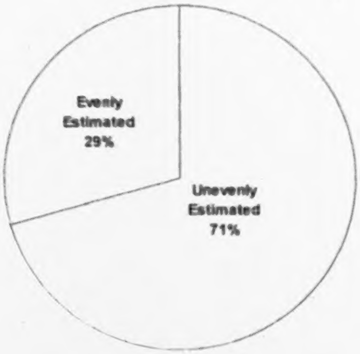


Figure 1.2: Variation between Evenly and Unevenly Distributed Cost Estimates

Despite the varied differences in the estimated and actual costs within a particular year, the City, being a public sector, must break even at the end of the year. Table 1.1 lists the difference between the actual and estimated costs sustained during the period of the study. The variation percentage shows how the City of Edmonton has normalized their charges to fall within the stipulated 10% that serves for budgeting from year to year.

Table 1.1: Total Project Cost during the Period of the Study*

Year	No. of Installations	Estimated Cost	Actual Cost	Difference	Variation
1999	57	**$4,345,230.55**	**$4,115,974.42**	**$229,256.12**	**5.57%**
2000	149	**$10,904,014.82**	**$10,339,699.16**	**$564,315.67**	**5.46%**
2001	157	**$12,855,115.61**	**$12,959,607.01**	**($104,491.39)**	**-0.81%**
2002	173	**$14,934,979.67**	**$14,095,151.94**	**$839,827.73**	**5.96%**
2003	136	**$13,843,904.48**	**$13,987,177.87**	**($143,273.39)**	**-1.02%**
2004	132	**$12,393,648.79**	**$12,557,294.31**	**($163,645.52)**	**-1.30%**

(* Note: Values in the table have been altered to protect the confidentiality of the information)

This thesis presents the methodology proposed to assist the organization's estimators in the estimating process. The proposed methodology has been incorporated into a computer system that integrates the developed Artificial Neural Network (ANN) model, IDMS, SAP and the SmartEST estimation software. This study analyses the data obtained from the current cost estimation procedure involved in the installation of water and sewer services. Based on the analysis of the data, various parameters have been utilized in the development of an ANN model to bridge the variation gap between the estimated and the actual costs. The data required for the development of this ANN model has been gathered from historical data obtained from the city's main-frame system, IDMS, which is used by the drainage services department.

1.2 Research Objectives

The main objectives of this research include:

1. Analyzing the activities involved in the current cost estimation practice of the City of Edmonton ;

2. Developing the ANN model and integrating the Temperature Forecasting Model *(Yu, H., et al., unpublished manuscript, 2005)* ;
3. Training the model based on the historical data obtained from IDMS and SAP;
4. Testing the Artificial Neural Network model.
5. Linking the developed ANN model with the existing SmartEST software.

1.3 Report Organization

Chapter 2 deals with the literature review, which looks into the aspects of the Artificial Neural Network and its significance to this research study. The chapter analyses the background of Neural Networks, the various types of networks, lists the application of Neural Network in the construction industry and, in particular, the use of Neural Networks for cost estimation in the construction industry. Chapter 3 details the current practices that the City of Edmonton has incorporated in the estimation procedure. Chapter 4 suggests a proposed methodology and analyses the various factors that are responsible for the variation in estimates. These factors later serve as inputs for the development of the Neural Network model, which has been detailed in Chapter 5 along with the selection of the network architecture. Chapter 6 details the observations that have been tabulated based on the selection of the network architectures. Chapter 7 describes the results of the project and also deals with suitable enhancement suggestions for the network, which serve as a scope for future work in this field.

Chapter 2: Literature Review

2.1 Introduction

The application and use of Artificial Neural Networks in construction cost estimation will be investigated in this chapter focusing on the following:

1. Background of Artificial Neural Networks
2. Types of Artificial Neural Network architecture
3. Learning and Training process of Artificial Neural Networks
4. ANN application in the construction industry
5. ANN application for construction cost estimation.

2.2 Background of Neural Networks

In 1956, the well known statement, "The potential use of computers and simulation in every aspect of learning and any other features of intelligence." *(Tsoulakas and Uhrig, 1997)* was defined at a conference at Dartmouth College. It was at this conference that the term `Artificial Intelligence' and `Neural Networks' gained potential. This was followed by the development of a model called the "Perceptron model"*(Rosenblatt, F. 1962)* that operated in the same manner as the brain. Minsky and Papert's 1969 book described the limitations of the Perceptron model which caused a setback in the interest in Neural Network research *(Minsky, M. and Papert, S., 1987)*. The development of algorithms, like back propagation, cognition and kohonen networks, in the early 80's resurrected the interest of the use of neural network technology. These algorithms gained potential through the nineties during which the back propagation network gained the most popularity. Artificial Neural Networks are being utilized in many commercial applications, such as character and image recognition, credit evaluations, fraud detection, insurance and stock forecasting *(Tsoulakas and Uhrig, 1997)*.

The Biological Analogy of the Firing Threshold of a Neuron: During the 60's and the 80's, the main branch of Artificial Intelligence research evolved rapidly to produce Expert Systems, which were based on a high-level model of logistic reasoning processes. Despite the application of these modulated systems in many domains, the developed systems still lacked the key aspects of human intelligence. This could be attributed to the fact that these systems are unable to replicate the pattern structure that the human brain structure is capable of handling.

The brain is composed of a large number of neurons that are massively interconnected. Each neuron is a specialized cell which consists of a cell body or soma, the dendrites, and the axon. The dendrites receive signals from the axons of other neurons; the dendrites conduct impulses toward the soma and the axon conducts impulses away from the soma as illustrated in Figure 2.1. When a neuron is activated, it fires an electrochemical signal along the axon which crosses the synapses to other neurons, which may in turn fire. However, a neuron fires only if the total signal received at the cell body from the dendrites exceeds a certain magnitude (intensity), which is called the firing threshold. The strength of the signal (information) received by the neuron critically depends on the effectiveness of the synapse. This concept was mapped into the new generations of ANN and its new technology [Web 1].

To accomplish complicated tasks, the brain involves a number of data processors (neurons) working in collaboration within a single unit. The Artificial Neural Networks have been programmed in the same manner.

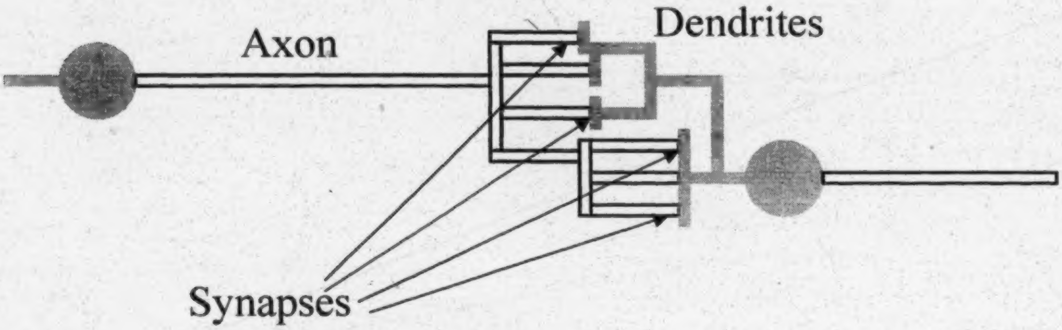


Figure 2.1: Information Transfer via a Synapse
[Web-2]

Artificial Neural Networks are simple electronic models based on the neural structure of the brain. It is a data modeling tool that is able to learn and retain complex input or output relationships. ANN models that are developed tend to resemble the human brain in the following two ways:

1. The learning process enables the Artificial Neural Network to acquire knowledge.
2. ANN knowledge is stored within inter-neuron connection strengths known as synaptic weights.

Figure 2.2 shows an illustration of a basic artificial neuron model simulating some basic functions. Neurons work by processing information provided to them through the inputs and utilize the synaptic weights associated with each input to provide resultant information in the form of spikes (signals).

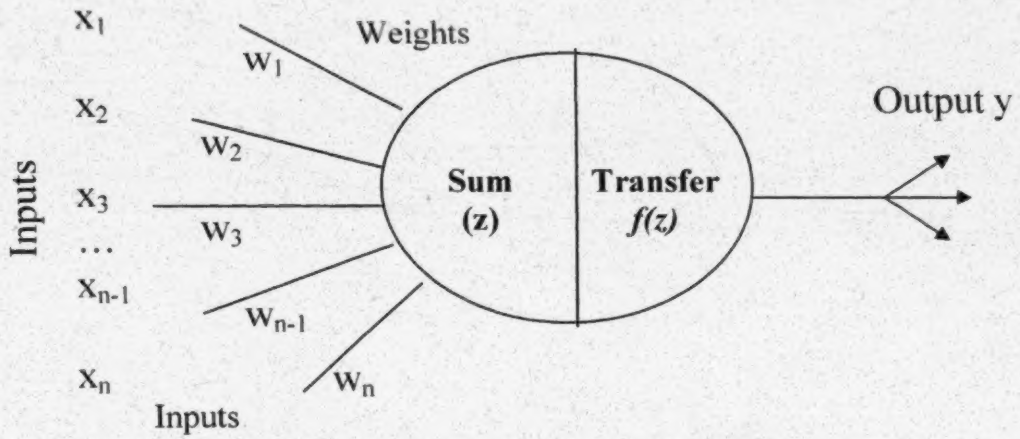


Figure 2.2: Basic Artificial Neuron (The McCullogh-Pitts Model)
(Tsoulakas & Uhrig, 1997)

The inputs to the network are represented mathematically by the symbol, x (n) which are in turn multiplied by a synaptic weight represented by w (n). The simplest case of a basic Neural Network involves the summation of a product of the inputs with their respective weights, and then fed through a transfer function to generate results, satisfying equations (1) and (2) *(Tsoulakas & Uhrig, 1997).*

$$\mathbf{z} = \sum \mathbf{x}_i \cdot \mathbf{w}_i \quad \text{(Sum of weighted Inputs)} \qquad \textbf{(1)}$$

$$\mathbf{y} = f(z) \quad \text{(Transfer) } \textbf{where y} \text{ is the neuron's output} \qquad \textbf{(2)}$$

Figure 2.2, along with the illustrations above, describe an individual neuron. However, for a network to be useful practically there has to be many such individual neurons, with variable inputs relative to the outside world, that have to be interconnected together to frame a definite model based on a particular form of architecture to produce an output. This serves as predictions or control signals. While all network models don't always relay to a single layer of inputs relating to an output, there can be hidden neurons that play an internal role within the network. This is the basis for the *Multilayer Perceptron* (MLP) theory or a *Feedforward Network* whereby the signals flow from inputs, forward through any hidden units, eventually reaching the output units *(Haykin,S., 1994)*. Figure 2.3 illustrates a schematic diagram of a typical network architecture following the MLP theory.

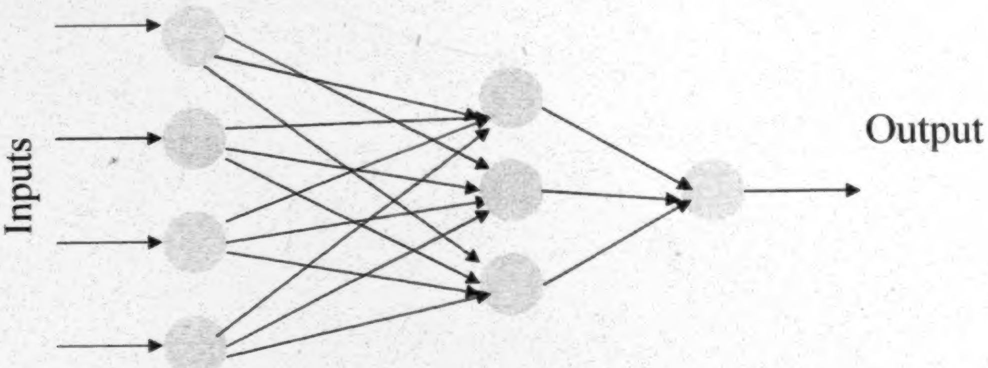


Figure 2.3: Typical Network Architecture
[Web-2]

In normal cases the functional output is a sigmoidal neuron, satisfying equation 3, but doesn't restrict itself to this and derives its function based on the architecture designed.

The sigmoidal function can be mathematically defined in equation (3):

$$s(z)= 1/ (1 + e^{-\alpha * z}) \qquad (3)$$

where:

α = coefficient that adjusts the abruptness of this function.

The value of α changes between two asymptotic values and is normally chosen between the range of 0.5 and 2. As a starting point α =1 can be utilized, and during the course of fine tuning the network, it could be adjusted within the given range. A graphic representation of the sigmoid function is shown in Figure 2.4 and is mathematically represented as:

$$s(0) = 0.5;\ \lim_{z \to \infty} s(z) = 1;\ \lim_{z \to -\infty} s(z) = 0$$

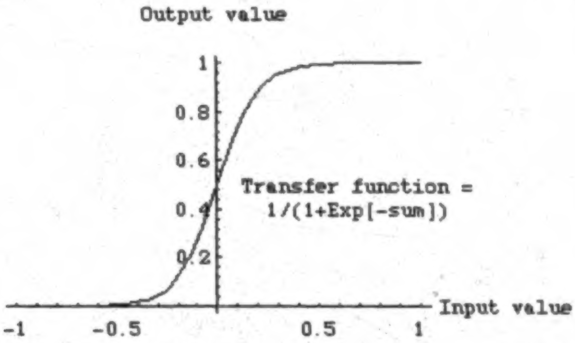


Figure 2.4 Sigmoid Transfer Function
(McMurrey,D., 2001)

$1/(1 + e^{-a*z})$ shuts down the path between the neurons if the signal strength is lower than 0.5, i.e., rounds down 0.5 to 0. If the signal strength is larger than 0.5, the function rounds it up to 1, hence opening the path to allow signal propagation.

Despite the relative advances in this field of study, it has been observed that the true power and advantage of Artificial Neural Networks lies in their ability to represent both linear and non-linear relationships and their ability to learn from these relationships. The conventional regression modeling is limited in the context of highly non linear systems. In an ideal setting it would be most appropriate to know the effective relationship between all the input parameters and their recognition with the output. However, it is difficult to understand the exact nature of the relationship between the inputs and the outputs within an ANN model due to the non linear nature of the data. As a result, the key feature is to enable the neural network to be in a position to learn the input-output relationship through the training process of the model.

The need for using Neural Networks: One of the most widely used predictive methodologies has been the regression methodology. Regression analysis has been used extensively to predict a continuous dependent output from a number of independent variables. Regression analysis is common among naturally-occurring variables, as opposed to experimentally manipulated variables. (*Tabachnick & Fidell, 1989).* Despite the wide applications of regression models, the major weakness is the need for a priori model that is chosen by the user. The data is then fitted to the model, regardless of how non linear in characteristic the data is. Neural Networks have the ability to obtain meaning from complex or imprecise data by their tendency to extract patterns and detect trends that are too complex to be noticed by either humans or other computing techniques. Once the Neural Network has been trained based on the category of information that it has been provided to analyze, it can be considered to be an `expert' in that field, which can offer projections given new situations. The main reasons for utilizing the ANN include:

1. The adaptive learning ability enables the ANN to understand the information based on the training provided.
2. The capability to self organize the data presented to the network and to detect their emergent contribution to the output.
3. Distributed Associated Mapping of Neurons: The data is uniformly distributed over many units providing resistance to noisy data, thereby

allowing the ANN to start with the noisy data and to recall the correct data.

4. Real Time Operation: ANN computations may be carried out using parallel algorithms, and special hardware devices are being designed and manufactured that take advantage of this capability.
5. Fault Tolerance via Redundant Information Coding: In any ANN model, the partial degradation or alterations of perceptrons leads to just a slight degradation in the behavior of the network as a whole.

2.3 Types of Artificial Neural Network Architecture

In general, Neural Networks are divided into two kinds: the ***heteroassociative*** Artificial Neural Network, in which the output vector is different from the input vector, and the ***autoassociative*** Neural Network, in which the output vector is identical to the input vector. *(Tsoulakas & Uhrig, 1997).*

Feed-forward Networks: These types of networks allow a one way travel of signals; from input to output (Figure 2.5). As a result, there is no loop (feedback), which means that the output layer of any network does not affect the same layer. Feed-forward networks are straight forward networks that associate inputs with outputs. They are extensively used in pattern recognition. This type of organization is also referred to as bottom-up or top-down.

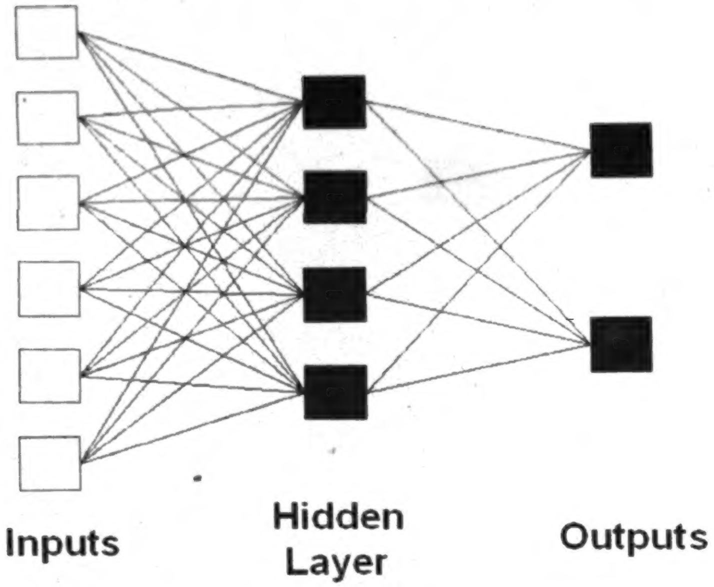


Figure 2.5: A Simple Feed-Forward Network
[Web-3]

Feedback Networks: These types of networks can have signals traveling in both directions by introducing loops within the network (Figure 2.6) and are dynamic in nature. Their 'state' changes continuously until they reach an optimal stage (sometimes this optimal point is a local minima as opposed to a global minima) corresponding to the lowest error. This stage is usually referred to as the equilibrium point. They continue to remain at the equilibrium until there is a change in the input vector, which results in the attainment of a new equilibrium. Such network architectures are also referred to as interactive or recurrent networks (e.g., Jordan-Elman networks), although the latter term is often used to denote feedback connections in single-layer organizations. These networks are very powerful and can be extremely complicated.

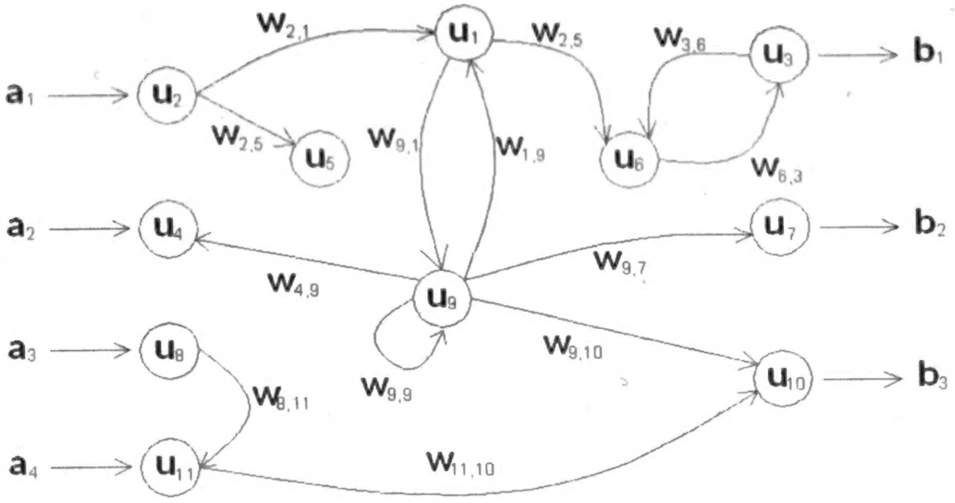

a1
a2
a3
a4
u1
u2
u3
u4
u5
u6
u7
u8
u9
u10
u11
b1
b2
b3
W2,1
W2,5
W2,5
W3,6
W6,3
W9,1
W1,9
W4,9
W9,7
W9,9
W9,10
W8,11
W11,10

Figure 2.6: A Feedback Network (Jordan – Elman type network)
[Web-3]

2.4 Learning and Training Process of Artificial Neural Networks

The key feature of the ANN is the ability to learn the input-output relationship through the training process defined by the user. Different types of network architectures use different training processes, but effectively the two types of training are as follows: the *supervised* and the *unsupervised* training, of which the supervised is the most common. Supervised learning is one that incorporates an external teacher that directs the manner in which the output unit ought to respond to the input parameters. The paradigms of a supervised learning include error-correction learning, reinforcement learning and stochastic learning. Figure 2.7 illustrates the schematic representation of a supervised learning system.

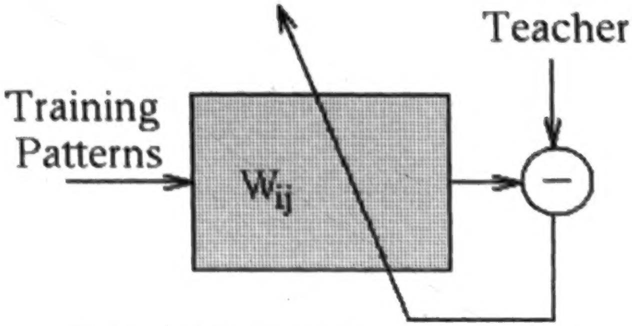


Figure 2.7: Supervised Learning System
[Web – 4]

The error convergence, i.e., the minimization of error between defined output values and the computed values from the model, has been an important issue concerning supervised learning. The main feature of the supervised learning system is the ability of the ANN to select a set of weights which minimize the error. The least mean square (LMS) convergence is to date the best-known methods of the learning paradigm.

In case of an unsupervised learning pattern, the training set consists only of input training patterns. Hence the network is trained without benefit of any teacher. The network learns to adapt based on the experiences collected through the previous training patterns. Frank Rosenblatt had coined the term perceptron in the 60's and defined the term as a connected network that simulates an associative memory *(Rosenblatt, F. 1962)*. Minsky and Papert (1969) showed the limitations of the single layer perceptron model of Rosenblatt resulted in the development of the Multilayer Perceptron (MLP) *(Minsky, M. and Papert, S., 1987)*. The training of such a perceptron depends on the architecture involved. This led to the development of a new algorithm within the Neural Network Architecture called the ***Back Propagation algorithm.*** This technique trains the network by calculating the error in the hidden layer weights through the error obtained in the output unit during the forward pass. For each data pair that is learned through a forward pass a backward pass is performed. This process is repeated over and over again until the error is at a low enough level *(Smith, L., 1996)*. Figure 2.8 shows a schematic representation of the back propagation network.

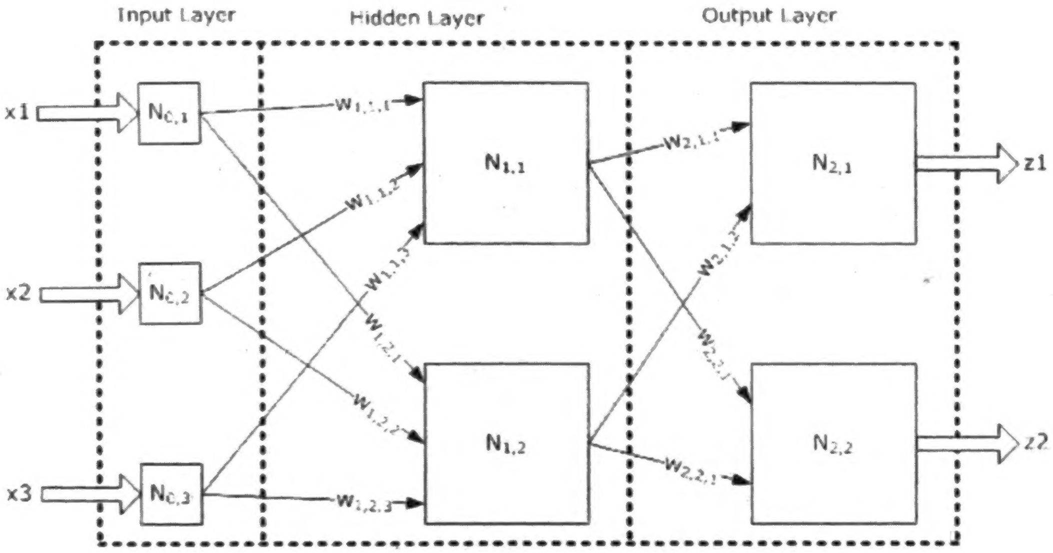

Input Layer
Hidden Layer
Output Layer
x1
x2
x3
N0,1
N0,2
N0,3
w1,1,1
w1,1,2
w1,1,3
w1,2,1
w1,2,2
w1,2,3
N1,1
N1,2
w2,1,1
w2,1,2
w2,2,1
w2,2,1
N2,1
N2,2
z1
z2

Figure 2.8 Back Propagation in Neural Networks
[Web-5]

The sample network consists of three different layers namely:

Input layer with three neurons; ***Hidden layer*** with two neurons; and an ***Output layer*** with two neurons. The following inferences can be associated with back propagation algorithms in Artificial Neural Networks. [Web-5]

a. The output of a neuron in a layer goes to all neurons in the following layer;
b. Each neuron has its own input weights;
c. The weights for the input layer are assumed to be one for each input value, i.e., the input values are not changed.

d. The output of the ANN is reached by applying input values to the input layer, passing the output of each neuron to the following layer as input.

e. The back propagation ANN must have at least an input layer and an output layer. However, it could have zero or more hidden layers.

The number of neurons in the input layer depends on the number of possible inputs that are available to the user, while the number of neurons in the output layer depends on the number of desired outputs. The number of hidden layers and how many neurons in each hidden layer can not be well defined in advance, and could change per network configuration and type of input data. In general, the addition of a hidden layer could allow the network to learn more complex patterns, but at the same time converges to an optimal state at a much slower rate.

2.5 ANN Application in the Construction Industry

The complexity of the structure and the application of ANN in the real business world, in particular the construction industry, have brought different levels of expectations which can be understood by the phrase: *"Neural Networks do not perform miracles. But if used sensibly they can produce some amazing results."* [Web–6] In this context, Artificial Neural Networks have been successfully applied in many industries and are gaining notice to their broad applicability in handling real world business problems. The ANN has been identified as a viable solution to identifying patterns or trends in data that can not be detected manually and are hence suited to predict or forecast needs in different fields like: a) sales forecasting; b) industrial process control; c) customer research;

d) data validation; e) risk management; and f) target marketing. The construction industry is a diverse industry with several general areas of specialization, some of which are related to resource allocation, scheduling, construction productivity and financial analysis. ANNs have been extensively used in both the commercial and industrial sectors.

ANN Application in estimating Construction Productivity: Chao and Skibniewski (1994) performed a study in which a Neural Network approach was developed in estimating the construction operation productivity of an excavator. Owing to the requirement for performing complex mapping of environmental (job conditions like soil conditions and characteristics of the excavator) and management factors (operation elements related to management of time related activities) with relation to the productivity, the Neural Network approach was best suited for this study. For the purpose of this study, two Neural Network modules were adopted, one for estimating excavator capacity based on job conditions; and the second for estimating excavator efficiency based on the attributes of operation elements. Based on the analysis of the data obtained, four main key factors were identified in the study which influenced the productivity, namely, the cycle time, the horizontal reach, the vertical type and the soil type. The output generated from the first network based on the cycle time data using an experimental desktop excavator is then incorporated into the second network, which examines the effect of operational elements on the productivity with the assistance of a simulation program to generate the production rate data. The success of this study lies in the fact that despite limited data collection effort, the proposed Neural Network model can produce a sufficiently accurate estimate thus having the potential to provide an efficient tool for construction productivity estimation.

The application of Artificial Neural Networks in industrial construction has found its influence in pipe handling, welding activities and estimation of construction productivity. One of the most prominent researches in this field has been associated with estimating construction labor productivity for concrete formwork tasks *(Portas, J. &. Abourizk, S., 1997)* In normal cases, the estimation of labor productivity is normally done through a combination of analytical techniques and personal judgment based on inputs from experienced personal, historical information and detailed work studies. Unlike normal Neural Network research employed within a feed forward back-propagation network resulting in a single output value for productivity prediction, this particular research involves using a combination of Neural Networks and fuzzy output layer to provide a frequency distribution histogram of the output value, reflecting the likelihood of the production rate. The reason for this uncertainty was mainly to due the fact that the single point estimate of the labor productivity was not acceptable to the estimators owing to the complexity of the problem. Moreover, those predictions that were wrong were found to be significantly wrong. The study also enlists a case study which incorporates the Neural Network model providing several benefits both academically and industrially which includes the ability to train inexperienced individuals by developing a structured approach to estimating construction productivity thereby reducing the guesswork associated with an estimate. The state of the art cited above was further developed on based on the development of a two stage Artificial Neural Network model that enables estimators to produce accurate labor production rates for industrial tasks like welding and pipe installation *(AbouRizk, S, et.al, 2001).* The data obtained for the purpose of this study was collected from a sample of 27 completed projects which involved 39 pipe installations. Based on literature reviews and surveys conducted , 33 different factors under 9 main categories were filtered out which include general project characteristics, site, labor, equipment, overall project difficulty, general activity conditions, quantity, design and activity difficulty. As compared to the earlier study by Portas and AbouRizk, this study was structured around one major difference. The data set was divided into two classes based on the type of production, namely, *"typical"* or *"non-typical"*, of which the typical multipliers are most encountered ones. This sort of classification provided two separate, yet harmonious, data sets where the use of a back propagation neural network would be ideal to train much easily than the entire set as a whole. The two stage network model first involved the classification of the network based on Kohonen's Learning Vector Quantization model. The values for the 39 records are preprocessed into a numeric format represented by an equivalent binary system. The second stage is the prediction of the Neural Network following the same output layer as discussed in Portas and AbouRizk utilizing the same input factors as that used in the classification network. The use of the two stage neural network helps reduce the error in pipe installation activity predictions in comparison to a simple back propagation network and in the long run reduces the subjectiveness attributed to an estimate from project to project.

The derivation of a probabilistic neural network classification model and its application in the construction industry was followed by an intermediate progress in estimating the labor productivity utilizing the Probability Inference Neural Network (PINN) model *(Lu, M et al, 2001)*. The model adapts the same concept as the model described by Knowles (2001) by incorporating the Kohonen's Learning Vector Quantization model with the combination of the probabilistic approach. The purpose for the use of this model was to enable estimators to make a decision for a future scenario based on the results recalled by the Neural Network model and personal preferences and experiences. Owing to this, the classification and prediction networks are combined to form an integrated network, which requires the development of a different training and recall algorithm. The PINN model uses a refined topology of the GRNN/PNN model which helps the model to generalize the statistical patterns within the training data and utilizing the iterative learning process to assist in coding these patterns into weight vectors following a four stage process. Three input data types are used to define the Neural Network input factors, namely Raw (quantitative input ratios), Rank (factors that are subjective in nature are converted into numeric format) and Binary (grouping of textual factors into numerical output). A Microsoft Access and Visual basic platform was developed to implement the training, testing, and recall for the PINN model. The PINN model was tested on real historical productivity data of 66 projects from a construction company resulting in a total of 81 input factors and validated and compared with a three layered feed-forward back propagation Neural Network. The application of the PINN model in industrial labor production rate estimating assists estimators to choose a course of action based on a better understanding of projects and the possible outcomes generated through the model.

The back-propagation neural network (BPNN) has been researched over the years and has been applied as a convenient decision-support tool in a variety of application areas in civil engineering industry. This study by Ming Lu (2001) specifies the sensitivity analysis of the BPNN by expressing the first order partial derivative between a Neural Network output variable and its input parameter. Learning algorithms such as the BPNN do not provide direct information on the effect of each input parameter or influencing variable upon the predicted output variable. Thus it is integral to test the response of a mature BPNN model to changes in sensitivity based on various input scenarios. An Artificial Neural Network such as the BPNN is merely an oversimplified representation of the real Neural Network in terms of structure and mechanism with the difference being that the BPNN has a multilayer structure each containing a number of interconnected processing elements between the layers. This study provides a mathematical term (both normalized and raw data) to define the input sensitivity of the BPNN algorithm in terms of the relationship between the output variable and an input parameter. An analogous comparison between the BPNN and regression analysis of statistics is discussed, and the sophistication and superiority of the BPNN over regression analysis is further demonstrated in a case study based on an artificial data set with four inputs, one output, one hidden layer with three hidden nodes and only ten records. A statistical analysis of input sensitivity based on Monte Carlo simulation was performed to further analysis the statistics of the first order partial derivative (slope) of the output signal over the input signal thereby enabling to understand the effectiveness of the BPNN model implementation in a probabilistic fashion. The sensitivity analysis of the BPNN is successfully applied to analyze the labor production rate of pipe spool fabrication in a real industrial setting. The data for this spool fabrication analysis is collected from 63 projects during the time period from 1995 to 1999 which involves data obtained from the material tracking system, weld tracking system, payroll system and questionnaires, surveys and discussions with various personnel's within the industry. Important aspects of the application, including problem definition, factor identification, data collection, and model testing based on real data, have been utilized as factors within the BPNN network.

One of the most ideal situations during any planning process is to be able to estimate the concrete placing productivity by providing an analytical model by changing the various parameters. The use of the neural network technology and in particular the feed forward and the Elman network architectures followed by the novel network architecture, *the Twin Nested Recurrent Network (TNRN)* was developed based on the concrete delivery process.*(Forbes, D et. al, 2004)* The factors were derived from previous studies done by previous studies that were included in this study including Portas and Abourizk's study *(1997)* and Rowings and Sonmez (1996) which deals with ten effects of multiple factors which simultaneously affect construction labor productivity likes quantities completed, humidity, precipitation, job type, number of workers, temperature, percentage labor, percentage overtime, cumulative properties and concrete pump. Rowings and Sonmez (1996) methodology integrates regression and neural network modeling techniques for quantitative evaluation of the impact of multiple factors on the productivity models for concrete pouring, formwork, and concrete finishing tasks, using data compiled from eight building projects. The TNRN architecture is unique to this study and involves the hybridization of a collection of standard networks connected in a novel manner providing a significantly promising result, which could find its use in other cyclic operations that provides not just a point prediction but one which provides a range of predictions for the various effects.

ANN Application in Structural Engineering: The application of ANN has developed a vast potential in the design of structures which involves a large number of variable parameters. The widespread application of ANN is in the structural optimization because of the capability to encapsulate the underlying behavior of structures. An example of the type of work done in this field is being done by the structural engineering research group at the University of Dundee, where a design tool based on the neural network technology is built to encompass current knowledge and the latest research conducted on the performance of cold-formed members *(El-Kassas, E. M. A., et al.).*

ANN Application in Lean Construction: ANN has found its way into lean construction principles. There are seven prerequisites defined for an activity's soundness in the construction industry, namely previous activities being completed, materials, information, equipment, labor, space and external conditions. An example of the operation of ANN in lean construction is described as follows. In any construction, an activity is considered as a node in highly connected feed forward network, and once this activity's inflows are fulfilled, the activity 'fires.' When completed by all, the seven requisites flow on to a new set of activities which may be within the same project, or in some cases, serve as inputs to subsidiary projects (*Koskela, L., 2000).*

2.6 ANN Application for Cost Estimation

The application of ANN for the purpose of cost estimations has found its use in complex projects. However, early stage cost estimation has still not been affected by these advances in comparison to most of the industry. It is still largely based on a combination of simple models and professional judgment; due to the fact that the aim of early stage estimation is to give the client a comparative cost analysis of "what if" scenarios. The added element involved in using a predictive approach, like ANN, is to provide the client with an objective forecast of the overall cost of the building. Such ANN models are developed to estimate the cost of projects by not just considering the physical characteristics attributed to a project, but also to model complex and little understood interrelationships that co-exist between factors that affect the cost.

Within the construction industry, the following aspects have related to the use of ANN as an appropriate cost estimation technique, namely: *(Duff, R. et al., 1998)*

a. ANN models, unlike linear regression, are able to model interdependencies between input data, which will inevitably occur when considering construction factors. For example, the model variables - such as number of storeys, gross floor area and number of lifts - will almost certainly be correlated.

b. ANN models can deal more readily with non-linear relationships that co-exist between cost related parameters.

c. ANN models can, more effectively than regression models, handle incomplete data sets, which is important since it is difficult to guarantee that complete data sets will always be available.

Despite the advances in the utilization of ANN as a key estimating tool, a project's final cost can only be forecasted to a limited certainty. The extent of the accuracy on a project's cost depends on the type of the project involved. The study of ANN has been employed using a stochastic Neural Network, which yields as its output not a single value but a distribution *(Harding, A. et al., 1999).*

ANN Application over Regression Methods: Cost estimation generally involves predicting labor, material, utilities or other costs over time given a small subset of factual data on "cost drivers" and statistical models, usually of the regression form, have assisted with this projection. The application of newer computational techniques, such as fuzzy logic and artificial neural networks, to the field of cost estimation have been considered in this study and the comparative study of the output results from the neural network and regression models for cost estimation have been carried out citing the case studies of costing of a pressure vessel by Brass and material cost estimation of carbon steel pipes by de la Garza and Rouhana. Brass claimed a 50% improvement when using a neural network instead of a regression model, however since the results hasn't had a sample to validate; it seems to be biased. This is known as the "re-substitution" method of model validation and is biased downwards. Despite these apparent faults, there have been significant improvements when using neural networks over the regression approaches. Shtub and Zimmerman in their study compared the costs of six product assembly strategies and found the neural network approach finer than the regression model. Smiths (1996) study is distinct from those just cited above by the completeness and integrity of the investigation that systematically includes the aspects of data set size, data set imperfections in the form of white noise and sampling bias, and the impact of model commitment in regression. This study examines the performance, stability and ease of cost estimation modeling using regression versus neural networks to develop Cost Estimating Relationships (CERs). The design of experiments tested four factors: the modeling method of developing the CER, the sample size available for CER construction, the magnitude and distribution of data imperfections (noise), and the bias of the sample. For each CER method, a full factorial experiment with five levels of construction sample size, three levels of noise and three levels of bias was created resulting in a total of 45 separate prediction models for each CER. The results generated from the comparative study show that the neural networks have advantages when dealing with data that does not adhere to the generally chosen low order polynomial forms, or data for which there is little *a priori* knowledge of the appropriate CER to be selected for the purpose of regression modeling. *(Smith, A.E, 1996)*

ANN Application in Parametric Cost Estimation: Owing to the limitations of the regression analysis a research study performed at the Cairo University looks at overcoming some of the drawbacks of Neural Networks. The research study presents a simple and effective approach for developing an adequate parametric cost models citing the gravity sewage projects as case study. The data required for the development of the neural networks is collected from various A/E firms from historical records about the sewage projects through the help of structured interviews and the distribution of surveys among professional designers. A simple neural network simulation has been developed in an excel spread sheet format and as alternative to Neural Network training, which was developed based on the commercial software, "Thinks-Pro", the Simplex Optimization and Genetic Algorithm were the two techniques used to determine the network weights. A comparative study between all the three patterns showed that Simplex optimization using Solver and the Neural Network Simulator provided the best results with reasonable errors. *(El-Gafy et.al, 2001)*

ANN Application in Cost Estimation in Structural Engineering: In terms of cost estimation, one of the studies done at Concordia University looked into development of neural networks in providing a detailed estimate for low-rise prefabricated structural steel buildings *(Siqueira, I, 1999)*. The data for the purpose of this study was obtained from a large manufacturer of prefabricated structural steel buildings in Canada featuring 75 building projects. The developed method employs neural networks for modeling individual project parameters associated with the direct cost of a project and aims at also improvising the cost estimate based on inputs such as cost adjustments, allocation of markups to individual cost items and taxes and the generation of reports. The proposed system generates the conceptual estimates at the planning stage so as to respond within the business market. The estimates developed at this stage provide reasonable accuracy in the estimation of direct cost of the buildings (material, labor, sub-contractor's cost, etc.,) assist in the decision making process and in the development of the project scope prior to bid submittals. In concurrence with the industry practice, three different neural network models were generated using the Neuroshell software namely, the order of magnitude (OM), parametric cost estimate for building walls (PW) and parametric cost estimates for structural steel framing (PS). In the development of each of the models, the GRNN approach out performed the back propagation models and was selected as best suited for this study. The validation of the model has been performed on data that had not been initially introduced into the training process and also through traditional regression analysis. The Automated Cost Estimate (ACE) software that is developed provides the conceptual estimate at the preliminary stage based on the neural network models, which is the advantage of this model.

ANN Application in Cost Estimation in Highway Construction: The application of Neural Networks has also found its way in the cost estimation of Highway Projects in developing countries. A research conducted within the Department of Civil and Engineering Department of the Saitama University, Japan, has enabled the incorporation of neural networks to provide a cost estimate with high accuracy at the conceptual phase of project development, which is crucial for planning and feasibility studies. *(Sodikov, J, 2005)*

Construction of such highway projects have a number of issues when conducting a cost estimate manually during the conceptual phase like the lack of preliminary information, lack of database of road works costs, missing data, lack of an appropriate cost estimation method, and the involvement of uncertainties. Given its significance, conventional tools such as regression analysis have been widely employed to tackle this problem but recent statistical studies show that errors in cost estimation have not decreased. Any project circle moves from the concept development phase, to the design, advertisement, bid/award, and finally the construction phase. As project progresses the accuracy of the cost estimation increases because project details becomes more clearly. At an initial stage, estimate accuracy is between about ±25% (minimum) and ±50% (maximum) *(Schexnayder, C. J., 2003)* due to less defined project details and other uncertainties due to both internal and external factors. The data required to develop the Artificial Neural Network Model, was obtained from the *ROad Costs Knowledge System (ROCKS)* developed by the World Bank Transport Unit to be used in developing countries. Thailand and Poland, the two countries that have the maximum number of projects were utilized as reference data for the development of the Artificial Neural Network model and sensitivity analysis were run based on pavement width, earthwork volume, work duration, average site cleaning and grubbing, surface class, and base material.

Chapter 3: Current Estimating Practice: The Case of the City of Edmonton

3.1 Introduction

This chapter describes the current practice adapted by the City of Edmonton for estimating project costs. The City of Edmonton's Drainage Design and Construction Section is responsible for the entire project, from project inception to project completion, and deals with all activities related to design, construction, and maintenance of the City's drainage infrastructure. This involves all activities ranging from the connection of water and sewer services to designing and building tunnels, pump stations, and storage treatment facilities.

The first step involved in the cost estimation for the water and sewer services is to determine the existing pipelines on site for a certain project. The as-built drawings for all projects are obtained from the IDMS database system and these drawings serve as the reference for the cost estimate. In certain cases, a site visit maybe required in order to initiate the development of a cost estimate. The estimator then performs a quantity take-off based on the requirement of the project, which serves as input parameters for the SmartEST. This is the current software utilized by the City of Edmonton in developing its cost estimate.

3.2 Estimation Software currently used

The SmartEST software is a MS Access based platform that has been developed by the Construction Engineering Management Group of the Department of Civil Engineering, University of Alberta, Edmonton, Canada and has been implemented successfully by the City of Edmonton. This particular software deals with estimation of jobs based on four different fields namely: labor, equipment, materials and other costs. The city adds certain percentages for overheads namely: construction overhead (13%), branch administration (7.5%) and warranty (0.5%). Since the data has to be manually fed into the system, the current SmartEST system adheres to the changes that could affect the variation in the value estimated and the actual cost. The parameters related to a particular task are stored in the database.

3.3 Current Estimation Procedure

The current process of estimation is based on the volume of material excavated; this approach is viewed as the best option viable when dealing with the projects of such nature. All the estimates are done on the basis of the four parameters namely: labor, materials, equipment and other costs.

Labor related costs: The labor is unionized and the payment is based on the contracts that normally don't change during the fiscal year. The crews are assigned to a specific type of job and work in correlation with each other.

Material related costs: Nearly all of the service material that is required for a particular job is available in stock within the City's storage unit for easy access. This particular procedure reduces the materials procurement time for jobs at hand. Most of the job tasks that require installations are mainly associated with copper pipes for water services and PVC piping for the sewer and drainage services. The fluctuation in costs is due to material pricing but is limited due to the incorporation of this procedure. The procurement of these materials from the City storage is distributed equally between the City officials and the foremen.

Equipment related costs: The equipment required for service installations depends on the type of job and the tasks. This may vary depending on the conditions on site that may arise, which are not taken into consideration while estimating a certain project. Depending on the complexity of a project, heavy equipments maybe required, which depends on the equipment availability over a given time frame.

Other costs: The other costs are related to barricading, surveys, saw cutting, coring, boring, asphalt paving or restoration based on the external factors that may arise on site. These details may or may not be available through the as-built drawings.

3.4 Results of Current Estimation Procedure

The City of Edmonton's current estimation technique has shown varied discrepancies between the estimated and actual costs during the period of 1999 to 2004. The data was obtained from the IDMS database and the results were plotted to show the degree of variation. Figure 3.1 illustrates an overall decree of projects that are within the (±10%) tolerance level set by the City management and the remaining that fall out of this range.

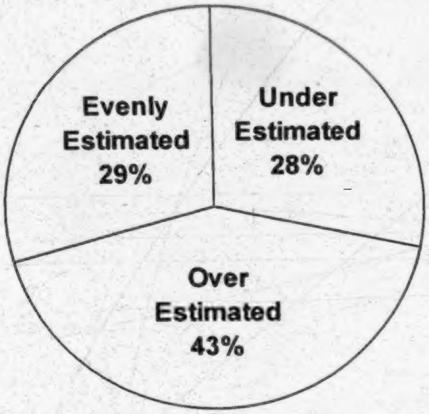


Figure 3.1: Service Installations Cost Percentage from July 1999 to Dec 2004

The study of the data obtained shows that only 29% of the total projects fall within the estimated tolerance limit of ± 10%. The remaining 71% of projects are not within this limit. Nearly 43% are over estimated while 28% are under estimated. An over estimated project refers to those projects whose estimated value is higher than the actual value incurred at the completion of the jobs. A detailed break down of all the services for 2004 has been shown in Figure 3.2 and Figure 3.3,, which illustrates that 75% of the projects fall out of the accepted ±10% of actual project costs. The data for the subsequent years from July 1999 till the end of December 2004 are plotted in Appendix A.

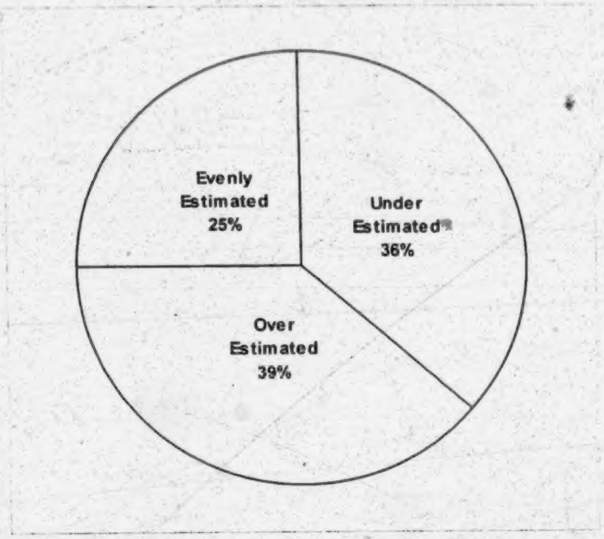


Figure 3.2: Estimated Costs for Service Installations- 2004

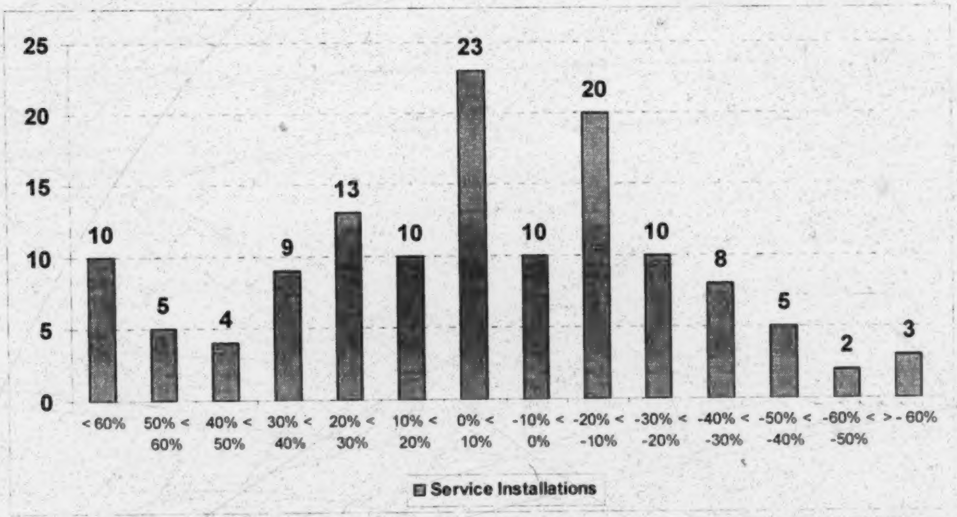


Figure 3.3: Number of Service Installations - 2004

The variation between the estimated and actual costs and the increase in the number of installations over the years from 1999 to 2004 have resulted in the need for an alternative approach by the City of Edmonton to improve the accuracy of their estimates.

Chapter 4: Proposed Methodology

4.1 Introduction

The City of Edmonton's Drainage Design and Construction Section is responsible for designing, constructing and maintaining the City's drainage infrastructure. This includes all activities from connecting water and sewer services to designing and building tunnels, pump stations and storage treatment facilities. Over the last six years, from the period of July 1999 to December 2004, there has been an increase in water installations of about 12%; 2002 having reached the peak, with over 170 installations. The large number of public works projects, coupled with a tight budget, has resulted in the need for the City of Edmonton to consider reviewing the overall productivity of its operations. Management within the City of Edmonton implemented an organization wide initiative to increase productivity within each of its departments.

This chapter discusses the proposed methodology that has been undertaken when dealing with the data obtained from the SmartEST and the analysis of the various factors required for the development of the Artificial Neural Network.

4.2 Proposed Research Methodology

The research methodology consisted of four phases. The first phase evaluated the current cost estimation procedure as implemented by the City of Edmonton using the SmartEST software. The results observed at this phase served as a baseline performance, which was used to assess the discrepancies associated with the estimation technique. The second phase involved assessing the work activities, from project inception to project completion. This phase is intended to seek alternative ways of providing a more defined structure while improving the cost associated with each work activity. This also involved detailing out features from the existing data obtained from IDMS and the output from this phase was used to provide a broader classification in determining the discrepancies associated with the baseline performance. The third phase of the project is technically oriented and utilizes the data obtained from the previous phases for the development of the ANN model using the Neuroshell® 2 ver 4.0 software. The model has been developed while accounting for the non quantifiable factors, such as the fluctuations in temperatures, which lead to the difference in the productivity with respect to standard costs for labor, materials and equipment. The selection of the input parameters was done through a series of interviews and discussions with personnel from various disciplines within the drainage services department. The goal for the ANN model is to obtain a reasonable confidence level in the prediction of the cost estimate for the installation of the water and sewer services. The last phase of the project involves the training of the model to perform in accordance with the set requirements. Graphical representation of the estimated costs and the actual costs incurred has been plotted. The model has been trained for data that is randomly distributed and considers the variation in costs over a given month in subsequent years. Figure 4.1 illustrates the focus of the methodology for the main process involved in the development of the Artificial Neural Network model.

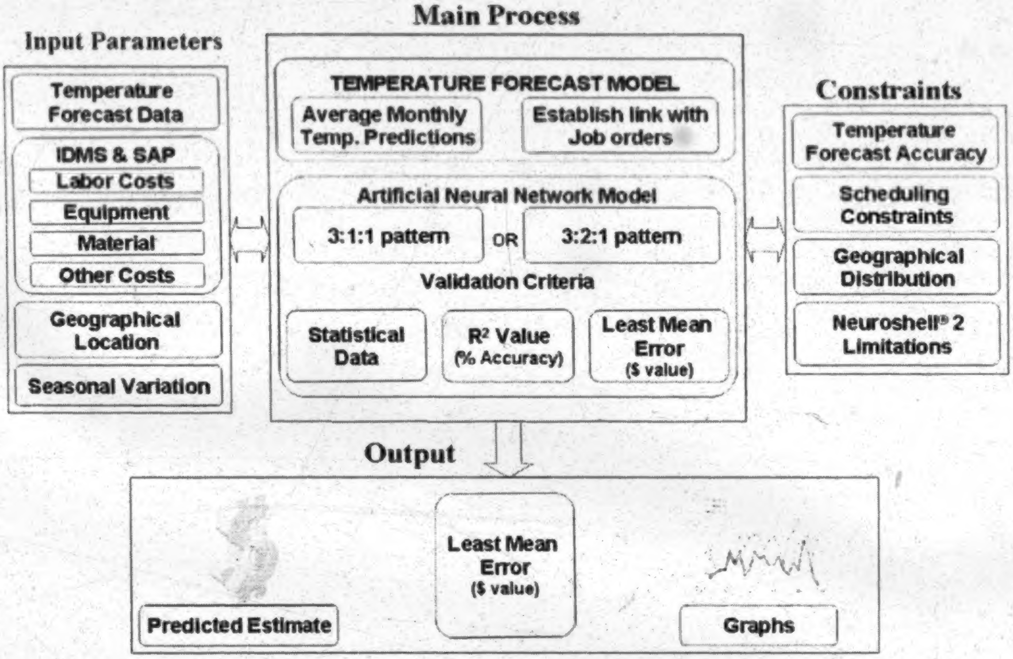

Input Parameters
Temperature Forecast Data
IDMS & SAP
Labor Costs
Equipment
Material
Other Costs
Geographical Location
Seasonal Variation
Main Process
TEMPERATURE FORECAST MODEL
Average Monthly Temp. Predictions
Establish link with Job orders
Artificial Neural Network Model
3:1:1 pattern
OR
3:2:1 pattern
Validation Criteria
Statistical Data
R² Value (% Accuracy)
Least Mean Error ($ value)
Constraints
Temperature Forecast Accuracy
Scheduling Constraints
Geographical Distribution
Neuroshell® 2 Limitations
Output
Predicted Estimate
Least Mean Error ($ value)
Graphs

Figure 4.1: Research Methodology Flow Chart

The figure relates to the overall procedure that the proposed study will undertake by incorporating factors such as the project estimation data from IDMS and SAP, geographical location of the project and project schedules as input parameters for the Neuroshell® 2 ver. 4.0 (ANN software) incorporating the average monthly temperature obtained from the Temperature Forecast model *(Yu, H., et al., unpublished manuscript, 2005).* The ANN model that is developed is expected to be integrated within SmartEST and used by the City of Edmonton as part of their real time operation.

4.3 Data Extraction

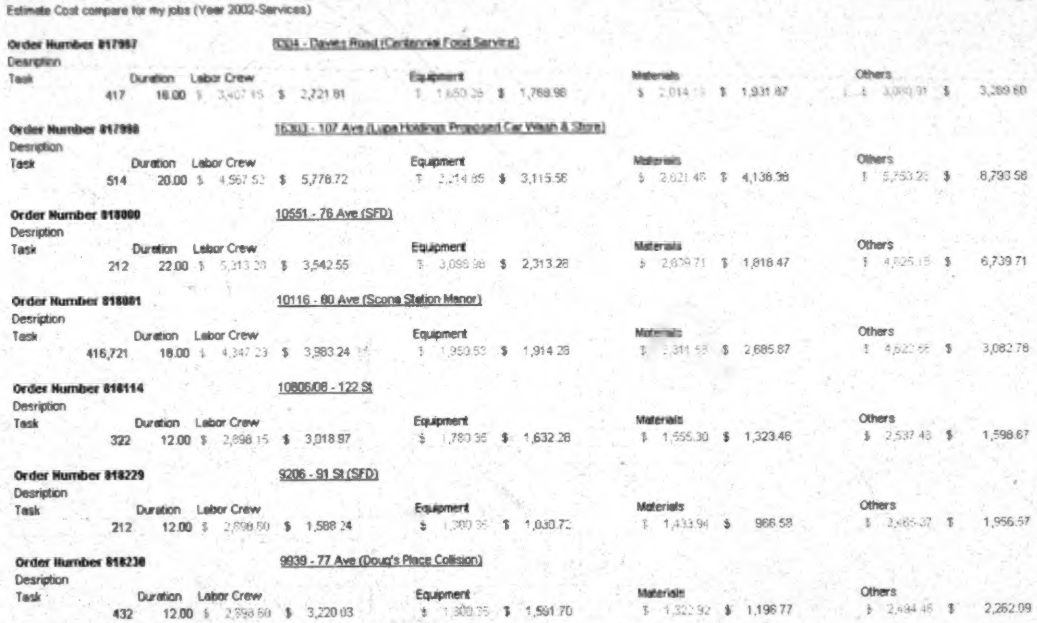
This is the first stage involved in determining the factors that contribute to the cost variation. In order to comprehend these factors more accurately, there was the need to analyze the data obtained from the IDMS mainframe about the related jobs. Table 4.1 shows the sample data of the job orders that have been extracted from the IDMS mainframe to the excel spreadsheets. A total of 804 projects are extracted over the period of the study (July 1999 to December 2004).

Table 4.1: Raw Data for a given job order

Job I D	Job Location	Task No.	Cost Elem.	Doc No.	document_date	description	ref_id	Type Hr.	Units	line_total	labour OH%	labour OH
816780	6633 - 118 Avenue	532	401000	220	03/12/2001	Peacocke, Rodney	393883	L	7	123.221	0.43	52.98503
816780	6633 - 118 Avenue	532	401000	220	03/12/2001	Fowler, Brian	483423	L	7	123.221	0.43	52.98503
816780	6633 - 118 Avenue	532	401000	114	03/12/2001	Nickerson, Brent	549978	L	8	144.12	0.43	61.9716
816780	6633 - 118 Avenue	532	401000	220	03/12/2001	Pringle, Gordon	317526	L	7	143.192	0.43	61.57256
816780	6633 - 118 Avenue	532	401000	220	03/12/2001	Kozinko, Henry	345345	L	7	138.285	0.43	59.46255
816780	6633 - 118 Avenue	532	401000	113B	03/12/2001	Habkirk, Brian	324601	L	7	143.192	0.43	61.57256
816780	6633 - 118 Avenue	532	401000	220	03/12/2001	Kobialko, Edward	115061	L	3	68.844	0.43	29.60292
816780	6633 - 118 Avenue	532	401000	220	/13/2001 00:00:0	Pringle, Gordon	317526	L	8	163.648	0.43	70.36864
816780	6633 - 118 Avenue	532	401000	220	/13/2001 00:00:0	Kobialko, Edward	115061	L	4	91.792	0.43	39.47056
816780	6633 - 118 Avenue	532	401000	220	/13/2001 00:00:0	Fowler, Brian	483423	L	8	140.824	0.43	60.55432
816780	6633 - 118 Avenue	532	401000	220	/13/2001 00:00:0	Kozinko, Henry	345345	L	8	158.04	0.43	67.9572
816780	6633 - 118 Avenue	532	401000	113C	/13/2001 00:00:0	Habkirk, Brian	324601	L	8	154.256	0.43	66.33008
816780	6633 - 118 Avenue	532	401000	220	/13/2001 00:00:0	Peacocke, Rodney	393883	L	8	152.112	0.43	65.40816
816780	6633 - 118 Avenue	532	401000	114	/13/2001 00:00:0	Nickerson, Brent	549978	L	5	90.075	0.43	38.73225
816780	6633 - 118 Avenue	532	401000	113G	/14/2001 00:00:0	Habkirk, Brian	324601	L	4	77.128	0.43	33.16504
816780	6633 - 118 Avenue	532	401000	220	/14/2001 00:00:0	Kozinko, Henry	345345	L	4	79.02	0.43	33.9786
816780	6633 - 118 Avenue	532	401000	220	/14/2001 00:00:0	Peacocke, Rodney	393883	L	4	76.056	0.43	32.70408
816780	6633 - 118 Avenue	532	401000	220	/14/2001 00:00:0	Fowler, Brian	483423	L	4	70.412	0.43	30.27716
816780	6633 - 118 Avenue	532	401000	220	/14/2001 00:00:0	Pringle, Gordon	317526	L	4	81.824	0.43	35.18432
816780	6633 - 118 Avenue	532	401000	220	/14/2001 00:00:0	Kobialko, Edward	115061	L	2	45.896	0.43	19.73528
816780	6633 - 118 Avenue	532	401000	114	/14/2001 00:00:0	Nickerson, Brent	549978	L	4	72.06	0.43	30.9858
816780	6633 - 118 Avenue	532	402000	220	/13/2001 00:00:0	Pringle, Gordon	317526	L	0.5	20.456	0.43	0
816780	6633 - 118 Avenue	532	402000	220	/14/2001 00:00:0	Pringle, Gordon	317526	L	0.5	20.456	0.43	0
816780	6633 - 118 Avenue	532	410004	220M	/14/2001 00:00:0	FILLCRETE 4178167		M	18	1260	0.43	0
816780	6633 - 118 Avenue	532	411000	123256	03/09/2001	ELBOW,PVC	67014	M	1	6.8452	0.43	0
816780	6633 - 118 Avenue	532	411000	123256	03/09/2001	COUPLING,PVC	54553	M	2	20.0342	0.43	0
816780	6633 - 118 Avenue	532	411000	123256	03/09/2001	PIPE,PVC	43472	M	9	186.6843	0.43	0
816780	6633 - 118 Avenue	532	411000	123256	03/09/2001	ADAPTER,PVC	61701	M	1	11.6521	0.43	0
816780	6633 - 118 Avenue	532	411000	123256	03/09/2001	ELBOW,PVC	61702	M	1	6.7471	0.43	0
816780	6633 - 118 Avenue	532	411000	123262	/13/2001 00:00:0	ELBOW,PVC	61702	M	3	20.2413	0.43	0
816780	6633 - 118 Avenue	532	411000	123262	/13/2001 00:00:0	CEMENT,GROUT	54781	M	3	26.9448	0.43	0
816780	6633 - 118 Avenue	532	411000	123262	/13/2001 00:00:0	ELBOW,PVC	67014	M	3	20.5356	0.43	0

Evaluating the data listed in Table 4.1 proved to be a difficult task in assessing the significant factors. As a result, there is the need to break the data to a more comprehendible and structured format that could easily facilitate the analysis of determining the external factors that contribute to the variation in the costs. Table 4.2 illustrates a sample of the data sorted out during a previous study *(Kung, D, 2005, unpublished M.Eng Report).*

Table 4.2: Data formatted during an earlier study

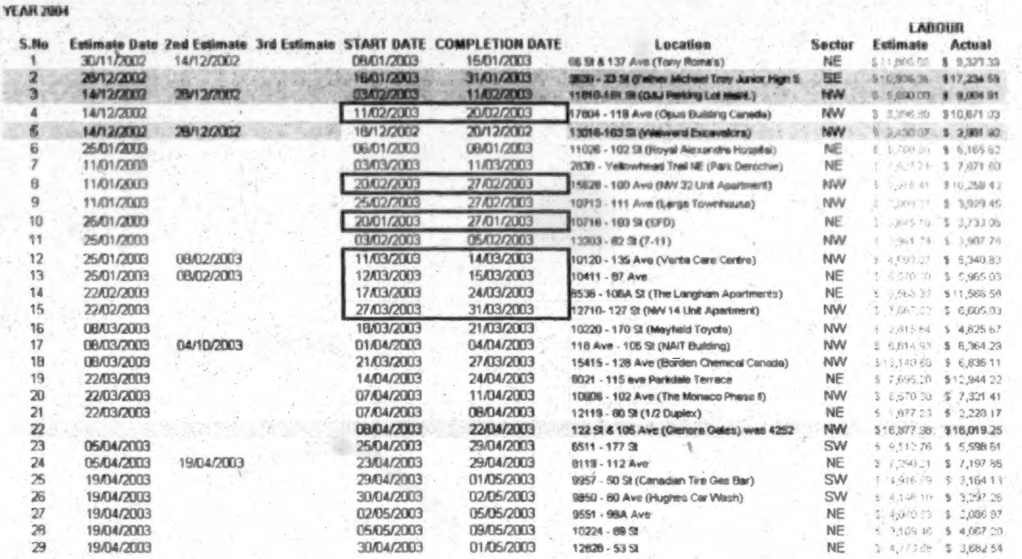


Estimate Cost compare for my jobs (Year 2002-Services)

Order Number	Desription	Task	Duration	Labor Crew		Equipment		Materials		Others	
817987	6304 - Davies Road (Centennial Food Service)	417	16.00	$ 3,407.15	$ 2,721.81	$ 1,650.26	$ 1,768.96	$ 2,014.15	$ 1,931.87	$ 3,090.91	$ 3,289.60
817998	16303 - 107 Ave (Lupa Holdings Proposed Car Wash & Store)	514	20.00	$ 4,567.52	$ 5,776.72	$ 2,214.85	$ 3,115.58	$ 2,621.46	$ 4,136.36	$ 5,753.23	$ 8,793.58
818000	10551 - 76 Ave (SFD)	212	22.00	$ 5,313.28	$ 3,542.55	$ 3,099.98	$ 2,313.28	$ 2,839.71	$ 1,818.47	$ 4,525.15	$ 6,739.71
818081	10116 - 80 Ave (Scona Station Manor)	416,721	18.00	$ 4,347.23	$ 3,983.24	$ 1,950.53	$ 1,914.28	$ 2,311.58	$ 2,685.87	$ 4,622.66	$ 3,082.78
818114	10806/08 - 122 St	322	12.00	$ 2,898.15	$ 3,018.97	$ 1,780.36	$ 1,632.28	$ 1,555.30	$ 1,323.46	$ 2,537.48	$ 1,598.67
818229	9206 - 91 St (SFD)	212	12.00	$ 2,898.50	$ 1,588.24	$ 1,300.36	$ 1,030.72	$ 1,433.94	$ 966.58	$ 2,465.37	$ 1,956.57
818238	9939 - 77 Ave (Doug's Place Collision)	432	12.00	$ 2,398.60	$ 3,220.03	$ 1,300.35	$ 1,591.70	$ 1,322.92	$ 1,196.77	$ 2,494.46	$ 2,262.09

The data listed in table 4.2 served as a baseline for the development of a more structured data as required by the ANN software, as seen in. table 4.3.The extracted data paved the way for understanding the projects in a more structured manner. It also enabled the extraction of valuable information that was required during the analysis stage for defining the relevant input parameters.

Table 4.3: Data formatted for easy Comprehension



YEAR 2004

S.No	Estimate Date	2nd Estimate	3rd Estimate	START DATE	COMPLETION DATE	Location	Sector	LABOUR	
								Estimate	Actual
1	30/11/2002	14/12/2002		08/01/2003	15/01/2003	66 St & 137 Ave (Tony Roma's)	NE	$11,866.00	$ 9,323.33
2	28/12/2002			16/01/2003	31/01/2003	3630 - 23 St (Father Michael Troy Junior High S	SE	$10,806.36	$17,234.59
3	14/12/2002	28/12/2002		03/02/2003	11/02/2003	11810-181 St (O&J Parking Lot Maint.)	NW	$ 5,690.00	$ 9,004.91
4	14/12/2002			11/02/2003	20/02/2003	17604 - 118 Ave (Opus Building Canada)	NW	$ 3,356.30	$10,871.03
5	14/12/2002	28/12/2002		18/12/2002	20/12/2002	13016-163 St (Welward Excavators)	NW	$ 3,430.07	$ 2,981.42
6	25/01/2003			06/01/2003	08/01/2003	11026 - 102 St (Royal Alexandra Hospital)	NE	$ 1,700.00	$ 6,165.62
7	11/01/2003			03/03/2003	11/03/2003	2630 - Yellowhead Trail NE (Park Derochie)	NE	$ 7,621.21	$ 7,071.60
8	11/01/2003			20/02/2003	27/02/2003	15828 - 100 Ave (NW 32 Unit Apartment)	NW	$ 9,918.41	$10,258.43
9	11/01/2003			25/02/2003	27/02/2003	10713 - 111 Ave (Large Townhouse)	NW	$ 2,909.77	$ 3,929.46
10	26/01/2003			20/01/2003	27/01/2003	10718 - 100 St (SFD)	NE	$ 3,845.10	$ 3,733.06
11	25/01/2003			03/02/2003	05/02/2003	13203 - 82 St (7-11)	NW	$ 3,941.79	$ 3,907.74
12	25/01/2003	08/02/2003		11/03/2003	14/03/2003	10120 - 135 Ave (Venta Care Centre)	NW	$ 4,597.07	$ 5,340.83
13	25/01/2003	08/02/2003		12/03/2003	15/03/2003	10411 - 87 Ave	NE	$ 6,570.00	$ 5,965.03
14	22/02/2003			17/03/2003	24/03/2003	8536 - 108A St (The Langham Apartments)	NE	$ 9,565.30	$11,586.58
15	22/02/2003			27/03/2003	31/03/2003	12710- 127 St (NW 14 Unit Apartment)	NW	$ 7,667.02	$ 6,605.03
16	08/03/2003			18/03/2003	21/03/2003	10220 - 170 St (Mayfield Toyota)	NW	$ 2,815.84	$ 4,825.67
17	08/03/2003	04/10/2003		01/04/2003	04/04/2003	118 Ave - 105 St (NAIT Building)	NW	$ 6,814.93	$ 6,364.23
18	08/03/2003			21/03/2003	27/03/2003	15415 - 128 Ave (Borden Chemical Canada)	NW	$13,140.60	$ 6,836.11
19	22/03/2003			14/04/2003	24/04/2003	8021 - 115 ave Parkdale Terrace	NE	$ 7,695.20	$12,944.22
20	22/03/2003			07/04/2003	11/04/2003	10806 - 102 Ave (The Monaco Phase II)	NW	$ 6,570.30	$ 7,321.41
21	22/03/2003			07/04/2003	08/04/2003	12119 - 80 St (1/2 Duplex)	NE	$ 1,877.23	$ 2,228.17
22				08/04/2003	22/04/2003	122 St & 105 Ave (Glenora Gates) was 4252	NW	$16,877.98	$18,019.25
23	05/04/2003			25/04/2003	29/04/2003	6511 - 177 St	SW	$ 9,112.76	$ 5,598.61
24	05/04/2003	19/04/2003		23/04/2003	29/04/2003	8118 - 112 Ave	NE	$ 7,290.21	$ 7,197.85
25	19/04/2003			29/04/2003	01/05/2003	9957 - 50 St (Canadian Tire Gas Bar)	SW	$ 4,916.79	$ 3,164.13
26	19/04/2003			30/04/2003	02/05/2003	9850 - 60 Ave (Hughes Car Wash)	SW	$ 4,146.10	$ 3,297.26
27	19/04/2003			02/05/2003	05/05/2003	9551 - 98A Ave	NE	$ 4,040.23	$ 2,036.97
28	19/04/2003			05/05/2003	09/05/2003	10224 - 89 St	NE	$ 3,189.46	$ 4,067.20
29	19/04/2003			30/04/2003	01/05/2003	12626 - 53 St	NE	$ 4,773.88	$ 3,882.54

4.4 Data Analysis

Data analysis proved to be the key structure to define the direction that the model should undergo while predicting the cost estimate. In order to obtain a more accurate prediction, input parameters to the network must be valid. The output, which in this case is the actual cost for each job order, enables the supervised learning of the network and makes the ANN control the parameters to adjust to the output. This section deals with the parameters that have influenced the cost variation between the actual and estimated costs. It has been observed from the formatted data that there has been a 12% increase in the service installations for the period of 1999 to 2004. The following parameters were found to show distinct variation in costs:

a) Amount of estimates performed per project

b) Seasonal Variation

c) Project Information such as duration between project estimation and completion

d) Geographical Location

***Amount of estimates performed per project**:* Most of the estimates are performed once. However, there are cases where the estimates are performed on more than one occasion and in some cases to the extent of being estimated three times. However, there are some cases where no estimate has been reported and this affects the credibility of the estimate. There are also some documented cases where the estimates occur either during the time the work was on site or after the work had been completed. It has been observed that despite the estimates being done after the job, it does not necessarily mean that the estimation is affected drastically by this factor, though it would definitely prove to be a contributing factor.

***Seasonal Variation**:* Since the service installations are done throughout the year, steps were taken to evaluate the seasonal affect on the cost estimation. The first approach split the year into the four seasons; this, however, didn't provide a clear indication on the seasonal affect. The adopted approach was to divide the year into the twelve months and to observe the monthly variation over the period of the study (July 1999 to December 2004). Figure 4.2 and Figure 4.3 illustrates a variation between estimates and actual costs of months within the given year (i.e., 2004).

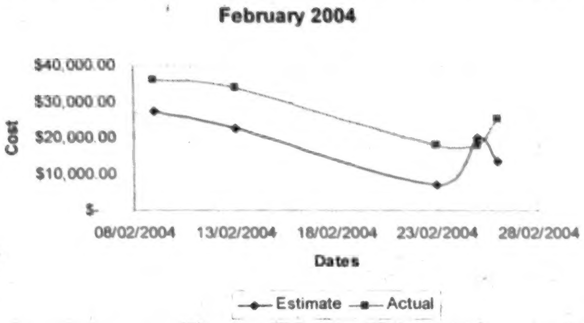


Figure 4.2: Variation in Estimated and Actual Costs – February 2004

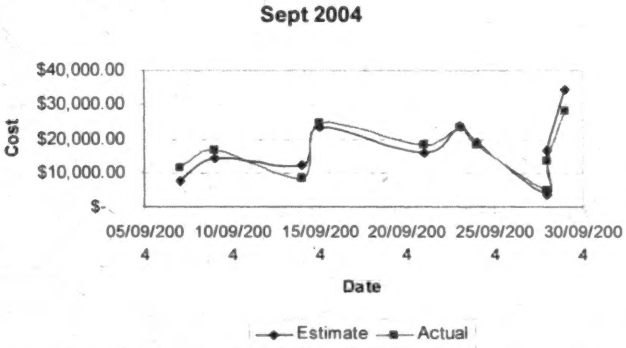
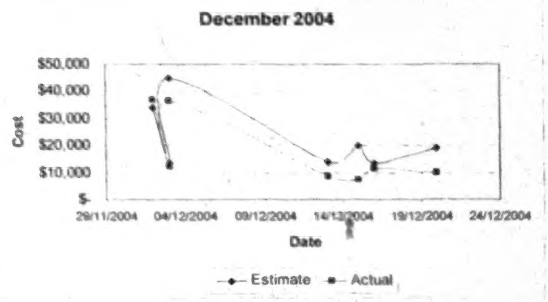


Figure 4.3: Variation in Estimated and Actual Costs – September 2004**
(** Note: Lines are drawn to enhance the readability of the graph)

A similar non regular trend was observed on all other months in the 6 years of the study period. Owing to the non regular aspect of this effect, the cost discrepancies between estimated and actual costs cannot be corrected by a constant factor. A comparative study was also performed between a given month over the 6 year period of the study to evaluate the effect of the months and the average mean temperature on the cost of the project. Figures 4.4 through till 4.10 illustrate this comparative study. Graphical representation of the various months over the 6 year period from July 1999 to the end of December 2004 can be found in Appendix B.



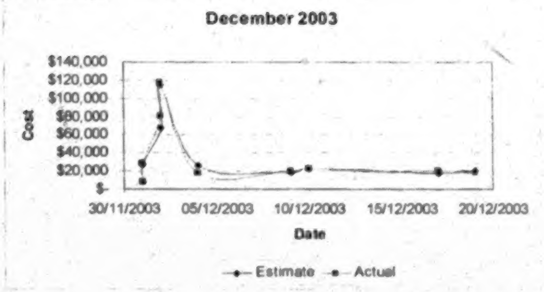
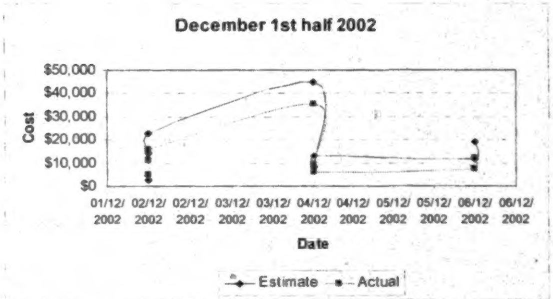


Figure 4.4: Cost Variation – December 2004

Figure 4.5: Cost Variation – December 2003



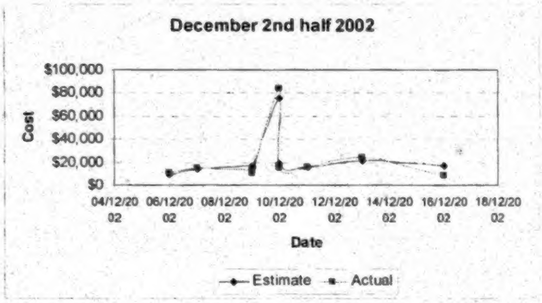
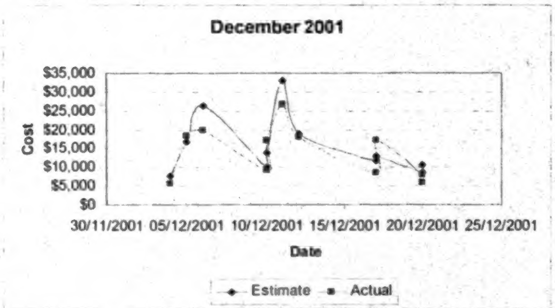


Figure 4.6: Cost variation – Dec 2002 (1st half)

Figure 4.7: Cost Variation – Dec 2001(2nd half)



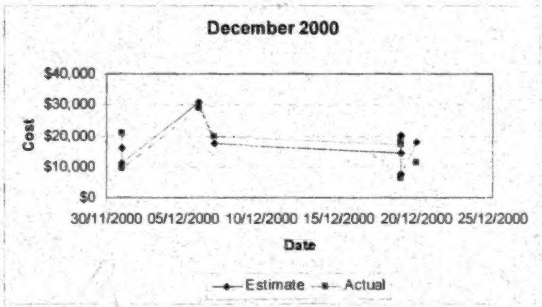
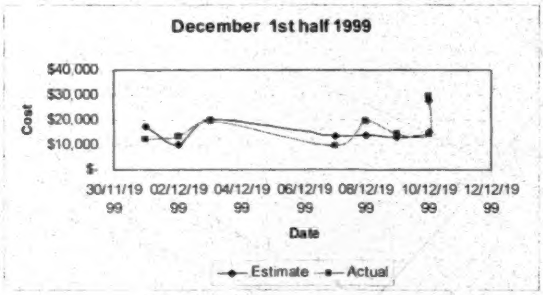


Figure 4.8: Cost Variation – December 2001

Figure 4.9: Cost Variation – December 2000



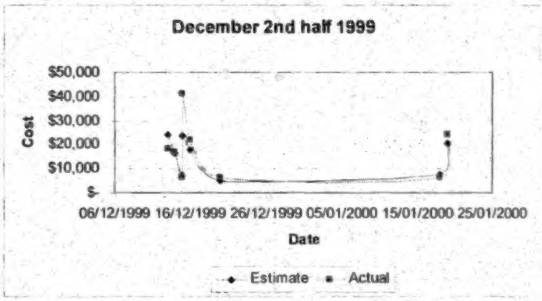


Figure 4.10: Cost Variation – Dec 1999 (1st Half)

Figure 4.11: Cost Variation – Dec 1999 (2nd Half)

The next step was to establish a link between the data obtained and the years involved in the study. The suggested attempt to facilitate this was to utilize the average monthly temperature over the years as an input parameter. For the purpose of this study, the Temperature Forecast model *(Yu, H., et al., unpublished manuscript, 2005)* was incorporated to predict the temperatures for the months ahead. This particular model was developed based on historical data obtained from the Edmonton weather network. Figure 4.12 shows the graphical variation of the monthly temperature of the years 1953 to 2003, which serves as the baseline for the Temperature Forecast model.

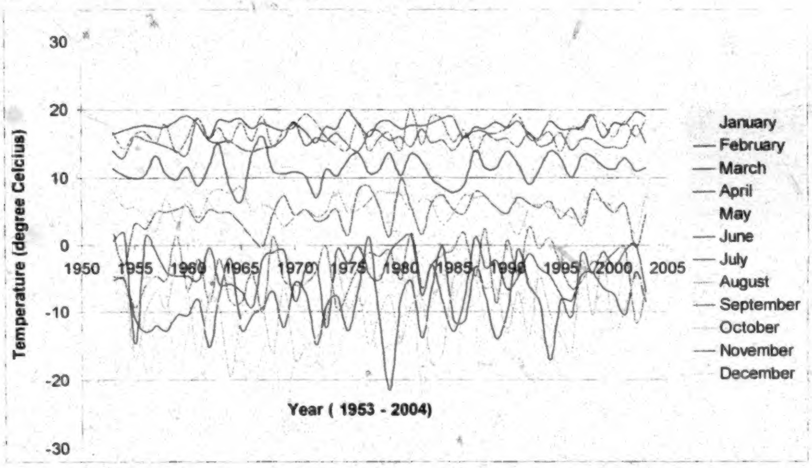


Figure 4.12: Monthly Weather from 1953 – 2003

On the basis of historical data obtained from the Edmonton weather network, the following prediction for 2005 was determined by the Temperature Forecast model, as shown in Table 4.4

Table 4.4 Forecasted Temperature for 2005 based on Yu, H., et al. 2005^{++}

Months	2001	2002	2003	2004	2005	2005
	Actual Mean Temperature	**AMT**	**AMT**	**AMT**	**Forecasted Temperature**	**AMT**
January	-2.7	-9.4	-11	-13.8	-11.4	-14.2
February	-10.5	-4	-8.5	-5.5	-7.4	-8.8
March	-0.4	-11.5	-6.7	-0.4	-2.3	-3.5
April	6	-0.4	4.7	6.3	5.3	5.7
May	12.8	10	10.6	9.4	11.3	10.4
June	14.5	17.7	15	15.4	15.4	13.8
July	17.8	19.6	19	18	17.6	15.5
August	18.3	16	18.5	15.9	17	13
September	12.8	10.9	11.4	10	11.9	8.9
October	4.3	1.8	7.4	3.5	4.7	4.4
November	-0.9	0	-7	0.7	-4	-1.3
December	-9.5	-4.7	-7.4	-7.8	-9.2	-6.9

($^{++}$ Actual Mean temperature data for 2005 obtained from Environment Canada [Web 7]

Project Information: This part deals with factors within a given project. These factors include the time taken between project estimation and completion; and the duration that is expected on a particular job. In most cases, the projects are completed on time, however, among the 804 projects, 33 projects have been found to be delayed. Also, there are tasks within a given job order that have been done after a couple of weeks from the date of the project completion, the longest one being 8 months. Another keen element of interest observed from the project information is the variation in costs based on the individual estimators, i.e., the human factor/effect. This was observed through analyzing the number of projects undertaken by each of the estimators and their proportion to the cost variation, which is illustrated in Figure 4.13.

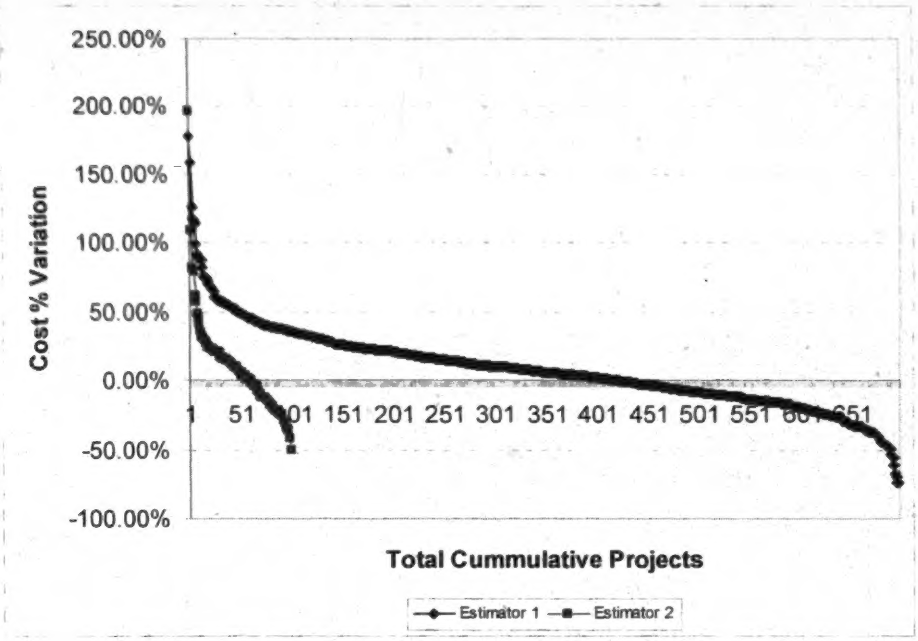


Figure 4.13: Cost Variation based on Estimators

The graphical representation in Figure 4.13 depicts the variation in estimates over the entire period of the historical data done by the two estimators. However, the main element to be considered is the number of projects undertaken by each of the estimators and their proportion with the cost variation. A better distribution of the cost variation by each estimator is illustrated in Figures 4.14 and 4.15.

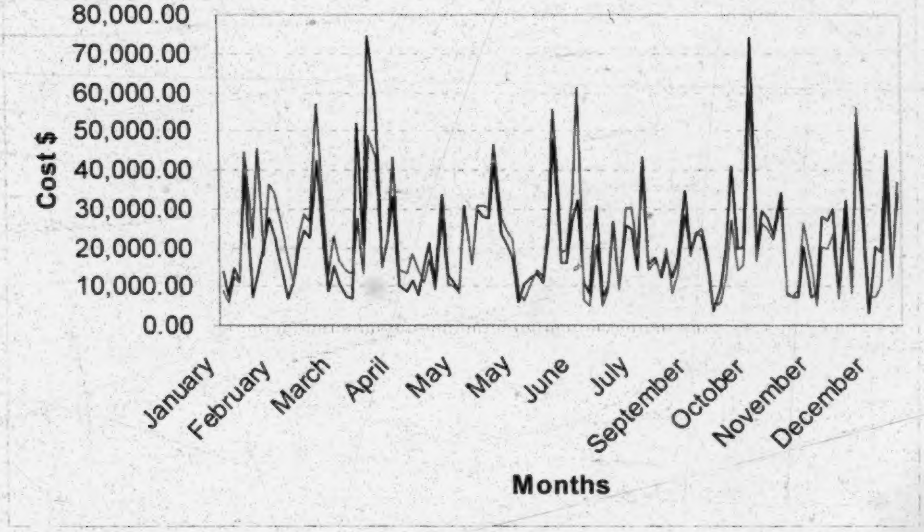


Figure 4.14: Cost Estimate by Estimator 1 – 2004

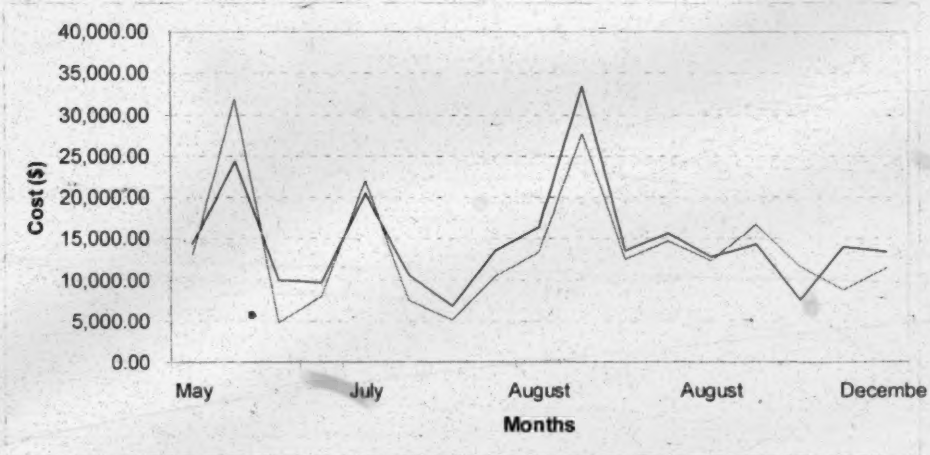


Figure 4.15: Cost Estimate by Estimator 2 – 2004

Geographical Location: Edmonton has been known to have different sand pockets in different areas within the city. As a result, the division of the jobs based on their geographical location was considered to be a major factor. Edmonton was divided into four major sectors; North West, North East, South East and South West. Initially, the City was divided into 5 sectors with the addition of Downtown. However, due to lack of sufficient data on projects from the downtown area, it was discarded. Figure 4.16 shows the four geographical sectors within the City of Edmonton; Whyte Avenue (82 Ave) divides the City between North/South and Calgary Trail (104 Street) divides the City between East/West [Web 8].

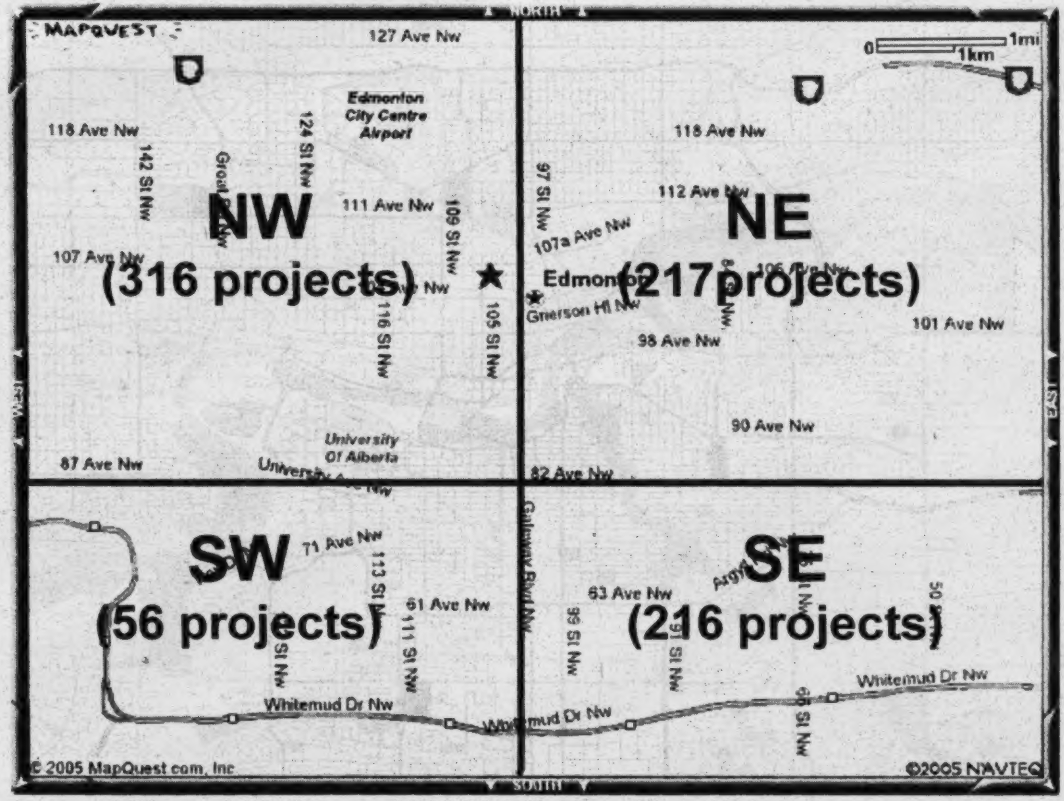

MAPQUEST
NORTH
127 Ave Nw
Edmonton City Centre Airport
118 Ave Nw
142 St Nw
124 St Nw
NW
(316 projects)
111 Ave Nw
109 St Nw
107 Ave Nw
116 St Nw
105 St Nw
97 St Nw
112 Ave Nw
107a Ave Nw
NE
(217projects)
Edmonton
Grierson Hl Nw
98 Ave Nw
101 Ave Nw
90 Ave Nw
University Of Alberta
87 Ave Nw
82 Ave Nw
0 1mi 1km
SW
(56 projects)
71 Ave Nw
113 St Nw
61 Ave Nw
111 St Nw
Whitemud Dr Nw
Gateway Blvd Nw
SE
(216 projects)
63 Ave Nw
99 St Nw
91 St Nw
66 St Nw
WEST
EAST
SOUTH
© 2005 MapQuest.com, Inc
©2005 NAVTEQ

Figure 4.16: Geographical Distribution: City of Edmonton

The variation in costs between estimated and actual costs for all the years from July 1999 to the end of December 2004 are plotted and are graphically shown in Figures 4.17 through 4.20. Graphs for the various geographical locations during each year are found in Appendix C.

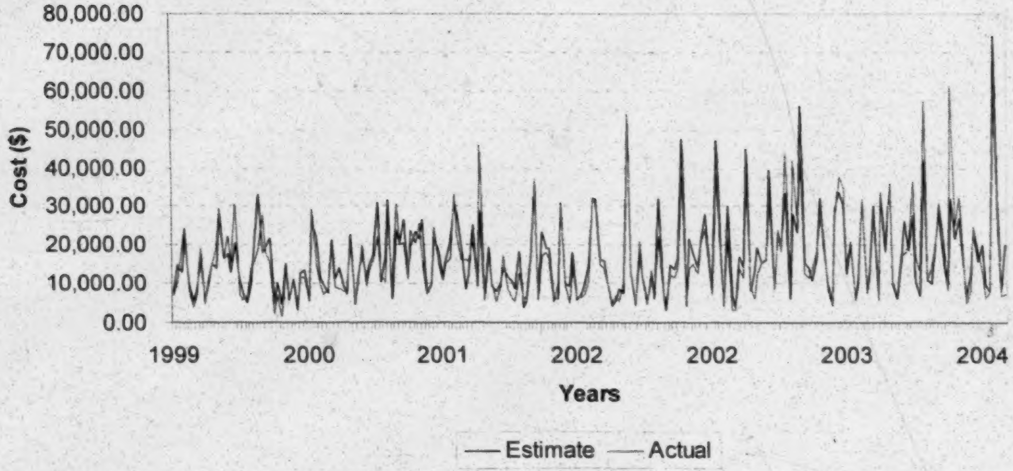


Figure 4.17: North East Sector

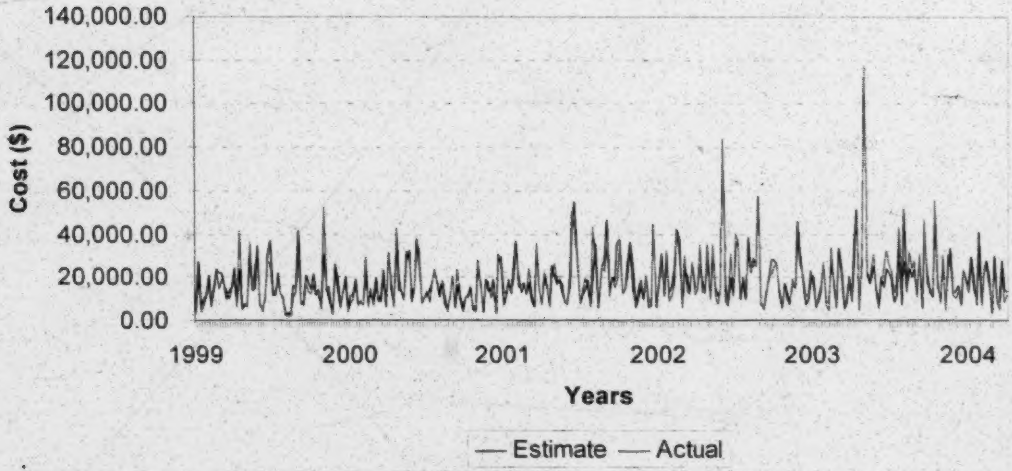


Figure 4.18: North West Sector

It is seen in Figure 4.16, that over the six year period of the study (1999 – 2004), most of the projects are within the North West and North East sectors. From Appendix C, it is noted that the North East sector of the city shows larger variation between the actual and estimate costs in comparison to the variation in the South West sector. However, this could be due to the fact that the number of projects within the South West sector is limited to 56 projects, in comparison to 217 projects installed in the North East sector.

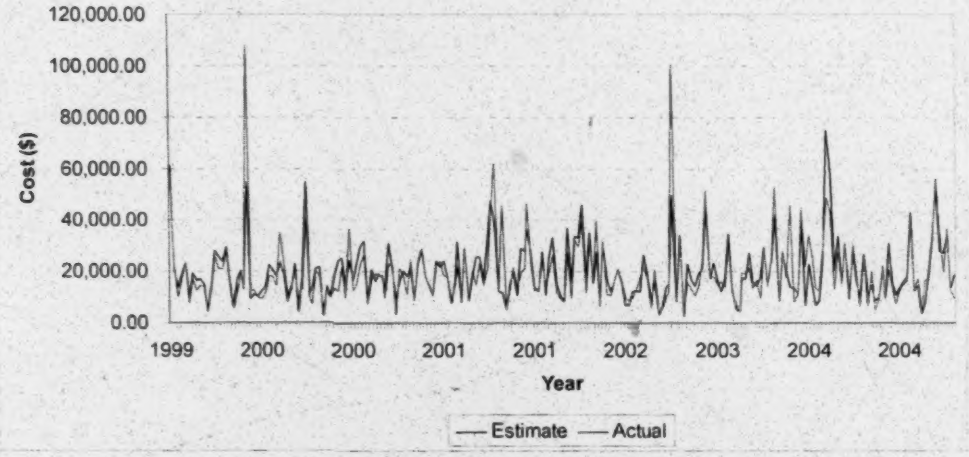


Figure 4.19: South East Sector

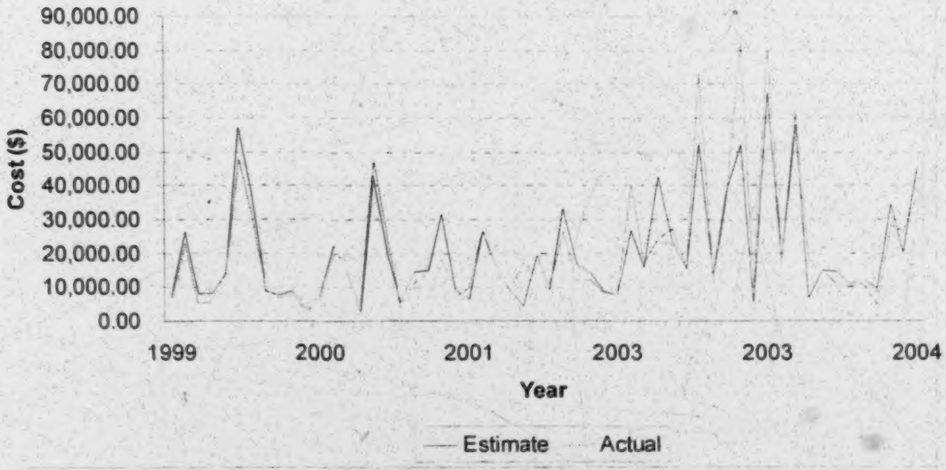


Figure 4.20: South West Sector

A geotechnical study would give a better understanding of the exact reasons of the variation shown in the figures above. The study could also channel many other geotechnical factors that could serve as influencing factors in the variation of actual and estimated costs.

Based on the analysis of the various data obtained from the IDMS, the following general factors were selected as possible input parameters for the development of the Artificial Neural Network model as illustrated in Table 4.5

Table 4.5 Neural Network Input Factors and Data Type

Input Factors	Data Type	Options and Remarks
Project Details		
Project Year	Binary	Grouped: (1999-2000), (2001-2002)
Project Month	Binary	Monthly distribution
Project Location	Binary	NW, NE, SW, SE sectors
Number of Estimates	Binary	1st , 2nd or 3rd estimate
Project Duration	Binary	Grouped: 2 days - based on monthly and weekly based analysis
Project Estimate Dates	Binary	Early/Late estimate
Project Status	Binary	In time / Delayed (option for forecasting in case of delay)
SmartEST details		
Labor Estimate	Raw	$ value, unionized , crew size
Equipment Estimate	Raw	$ value, hired , additional requirements
Material Estimate	Raw	$ value, additional, filcrete
Other Estimates	Raw	$ value, site conditions, change orders
Other Factors		
Human Factor	Binary	Estimator 1/ Estimator 2
Average Mean Temperature	Binary	Grouped: (14-16),(16-18),(18-20)....in Fahrenheit
Site Conditions	Binary	Barricading, Coring, Saw Cutting

4.5 Modeling Objectives

This section describes the steps following the analysis of the data obtained from the IDMS main frame system. The analysis of the data assists in obtaining the input parameters that are required for the development of the ANN model. The modeling objectives for the development of the Artificial Neural Network follow the process illustrated in Figure 4.21.

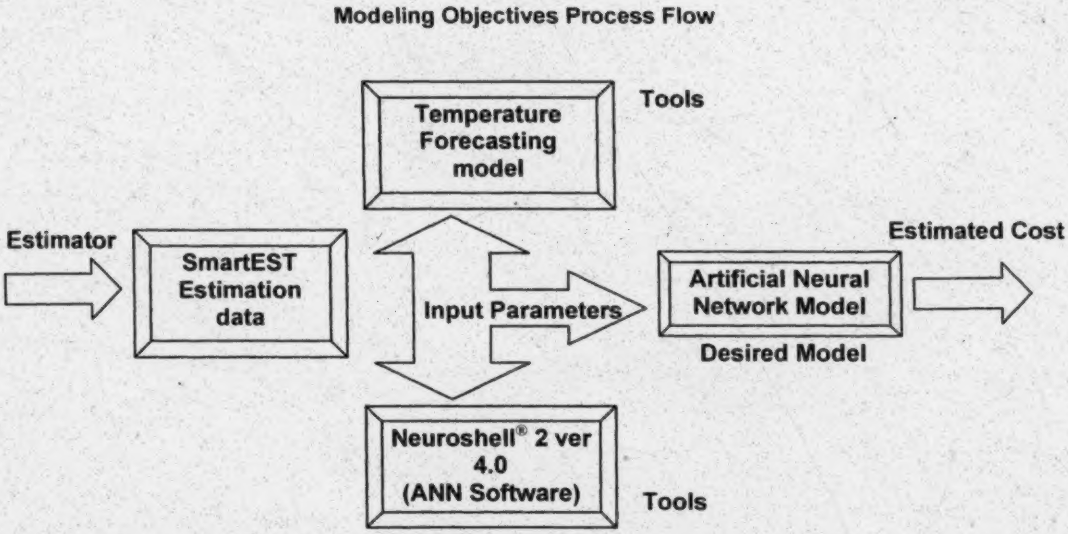

Modeling Objectives Process Flow
Temperature Forecasting model
Tools
Estimator
SmartEST Estimation data
Input Parameters
Artificial Neural Network Model
Desired Model
Estimated Cost
Neuroshell® 2 ver 4.0 (ANN Software)
Tools

Figure 4.21: Modeling Objectives Process Flow Diagram

The above figure explains the process flow for the development of the Artificial Neural Network model. The input parameters are obtained from the SmartEST estimation historical data after an extensive sets of interviews and discussions with the estimators and other personnel from the City of Edmonton. The average monthly temperature is predicted by the Temperature Forecasting model *(Yu, H., et al., unpublished manuscript, 2005),* which, along with other factors, serve as input factors for the ANN model utilizing the Neuroshell® ver 2.0 software. The output generated is the desired ANN model which predicts a better estimate. The main aim of the development of the ANN model is to link it with the current estimation software to assist the City estimators to provide an accurate estimate. Based on the analysis of the various factors described, and the estimate obtained from the SmartEST, an accurate predicted estimate is determined utilizing the Temperature Forecasting model and the Neuroshell® 2 ver. 4.0..

Chapter 5: Development of the ANN Model

5.1 Why use Artificial Neural Networks?

This section deals with the reasons for choosing Artificial Neural Networks as a solution to bridge the gap between the estimated and actual costs that the City of Edmonton has been incurring over the last six years. Considering the credibility and the nature of the data obtained, there were several factors leading to the decision to promote an Artificial Neural Network methodology among other methods. The factors that motivate the application of an ANN model include:

- On the basis of the data obtained and analyzed, it was difficult to determine a definite pattern between the various parameters by human or simple regression models.
- A rich literature documenting case studies similar to the work undertaken in this project.
- Separating the data into classes, namely for training, testing and validation, provided a baseline to allow the network to undergo an adaptive learning process thereby enabling quality assurance.
- Due to the raw nature of the data obtained from the IDMS mainframe, the data has to be organized before it is used.
- Since the proposed methodology incorporates the use of a predictive temperature model, the ANN would smoothen out any noisy data.

- The City of Edmonton's estimators are familiar with the Artificial Neural Network; in addition the ability to link the ANN with the SmartEST software motivates the use of this model.

5.2 Defining the Supervised Network

The first step in developing the network is to define the input factors on the excel spreadsheet that has been imported to the Neuroshell® 2. ver 4.0 software. Since this project deals with obtaining a certain output value, it is necessary to define an actual output column which assists during the supervised learning. The input factors include the years from 1999 to 2004, the amount of estimates performed, the time the estimates are performed (before, during or after work is completed), average monthly temperatures (1999-2004 Temperature Forecasting model), month within a year, project completion status i.e., delayed or not, geographical location, labor estimates, equipment estimates, material estimates and other costs. The input parameters are subjective in nature and cannot be easily quantified. The input parameters have been assorted to form a binary format (1,0) where 1 would accept the value and 0 would not accept the value.

Once the input factors are defined, the ANN software computes the maximums and minimums, along with the mean and standard deviation under each parameter. Artificial Neural Network models require variables to be scaled into the range of 0 to 1 or -1 to 1. This assists the network to determine the variable's real value range *(NeuroShell® 2 ,© Ward Systems Group, Inc.).*

5.3 Data Extraction

The next stage is to extract the test data. In this particular case, the extraction is done on the basis of classes. The 804 cases have been divided into three different classes that include training, testing and production/validation sets. These classes have been set in the ratio of 3:2:1 respectively. An initial ratio of 3:1:1 was considered. However, partitioning the data in this ratio showed a larger discrepancy in the R^2 value between the test data sets (62.05%) and validation data sets (83.05%). Chapter 6 deals with the various observations with regards to runs performed with both patterns and the different architectures.

The current pattern involves setting up a baseline reference, which is the percentage variation between actual and estimated costs, and has been set to 0% variation in order to minimize the bias in the training data sets for both over and under estimated projects. The patterns are organized based on the months in a year and attempts have been made to evenly distribute among the years by mirroring across the set baseline. The data retrieved 404 training sets, 270 test sets and 130 production sets based on the 3:2:1 pattern. Figure 5.1 shows the pattern distribution as described above.

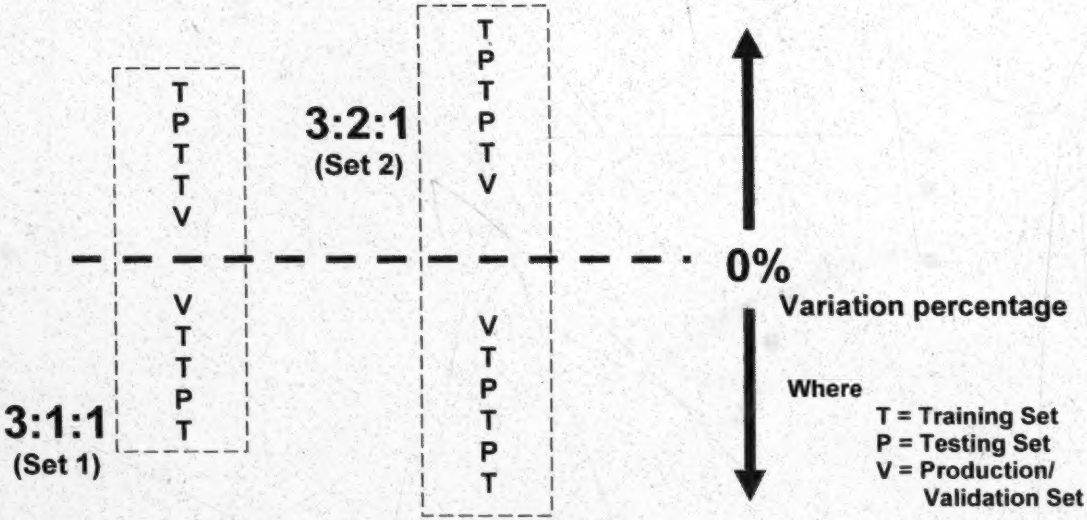
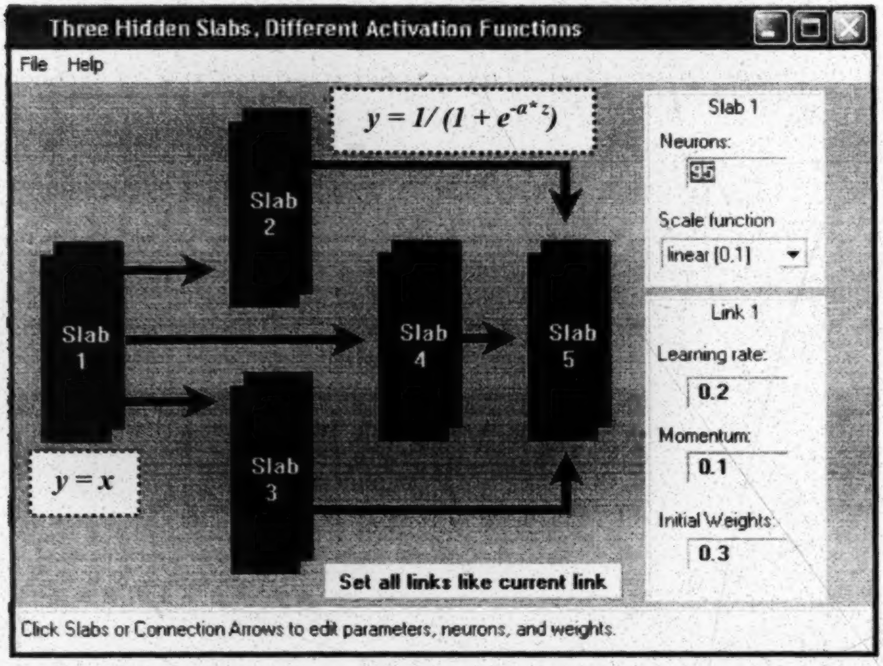

T
P
T
T
V
3:2:1
(Set 2)
T
P
T
P
T
V
0%
Variation percentage
V
T
T
P
T
3:1:1
(Set 1)
V
T
P
T
P
T
Where
T = Training Set
P = Testing Set
V = Production/
Validation Set

Figure 5.1: Pattern Distribution

5.5 Network Architecture Selection

The next step in the process is to set up the network architecture. Various types of network architectures where experimented on a trial and error basis. The ***"Jordan-Elmans network***" produced an output data that went through another filtering process and worked well with the 3:1:1 pattern. In the case of the 3:2:1 pattern, a ***"Ward network"*** with three hidden slabs (layers) is considered and is illustrated in figure 5.2. The first slab (layer) is independently connected to each of the hidden slabs (layers) following multiple activation functions to provide an output slab (layer). However, in this network architecture the same logistic function was used within the three hidden layers of the network architecture. After experimenting with networks it was found that the 3:2:1 classification produces the best results by combining stability and accuracy. The success of the prediction is attributed to the consistency between the R^2 values obtained in all sets (training, testing and production/validation data sets) and the least mean error value, which is the dollar value. Figure 5.2 shows the type of ANN architecture selected for this project.


Three Hidden Slabs, Different Activation Functions
File
Help
$y = 1/(1 + e^{-a*z})$
Slab 2
Slab 1
Slab 4
Slab 5
Slab 3
$y = x$
Set all links like current link

Slab 1

Neurons: 95

Scale function: linear [0,1]

Link 1

Learning rate: 0.2

Momentum: 0.1

Initial Weights: 0.3

Click Slabs or Connection Arrows to edit parameters, neurons, and weights.

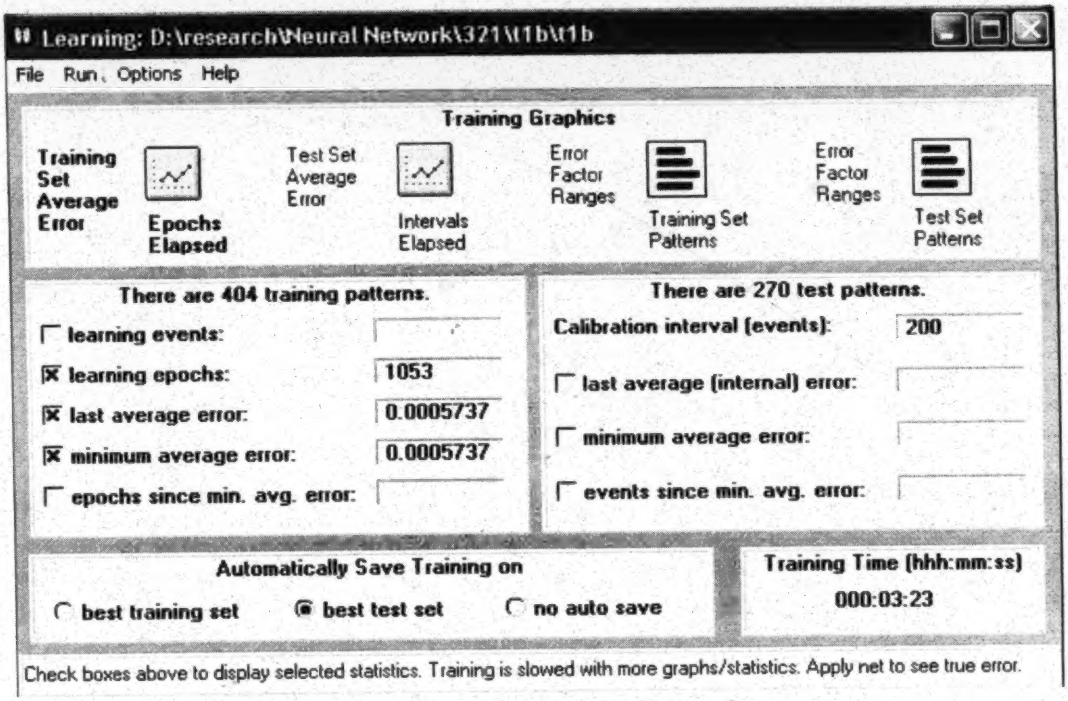
Figure 5.2: Ward Net – Three Hidden Slabs (layer): Network Architecture

The first slab (layer) of this network has 95 neurons following a linear function (0, 1) $(y = x)$ since the input needs to pass to the subsequent layers unmodified. A learning rate of 0.2 and a momentum of 0.1, with the weights initially set at 0.3, are used for all the slabs (layers). The momentum is the factor that helps proportionate the change in a previous weight to a new weight. This allows the model to converge faster to a solution that is optimum by reducing the oscillations of the weights which commonly occur in any non linear optimization procedure. The second, third and fourth slab (hidden layer) each fields 23 neurons and follows a logistic sigmoid function $(y = 1/(1 + e^{-a*z}))$. The final slab (layer) is the output slab (layer), which yields a single value and follows the logistic sigmoidal activation function. The selection of the network to yield the desired response is the supervised learning technique.

To train the network the following parameters were used within Neuroshell® 2 ver. 4.0 (*NeuroShell 2, © Ward Systems Group, Inc.)*.

1) ***Calibration level of 200:*** The value corresponds to a step at which the records are extracted from the data sets for the training process. For instance, after randomly extracting the first data, say R_k, the next record will be R_{k+200}. The process goes through the data as many times as necessary until the training data is exhausted.
2) ***Epochs set to 300,000:*** An epoch corresponds to a single pass through the entire set of training data. With this particular setting, the software goes through the data 300,000 times. For each pass the records are extracted randomly and a new set of parameters are obtained. The best set of control parameters are returned at the end of the process.

The learning process of the network is accomplished when the error is optimal for the best test average. Figure 5.3 depicts the learning process that the network went through.


File
Run
Options
Help
Training Graphics
Training Set Average Error
Epochs Elapsed
Test Set Average Error
Intervals Elapsed
Error Factor Ranges
Training Set Patterns
Error Factor Ranges
Test Set Patterns

Learning: D:\research\Neural Network\321\t1b\t1b

There are 404 training patterns.

☐ learning events:

☒ learning epochs: 1053

☒ last average error: 0.0005737

☒ minimum average error: 0.0005737

☐ epochs since min. avg. error:

There are 270 test patterns.

Calibration interval (events): 200

☐ last average (internal) error:

☐ minimum average error:

☐ events since min. avg. error:

Automatically Save Training on

○ best training set ● best test set ○ no auto save

Training Time (hhh:mm:ss)

000:03:23

Check boxes above to display selected statistics. Training is slowed with more graphs/statistics. Apply net to see true error.

Figure 5.3 Learning Process of the Network

On the basis of the epochs and the intervals elapsed, the training and test set average errors are graphically interpreted from the learning that the software undergoes. Figure 5.4 shows the optimization procedure that the network went through during the learning process to predict a more accurate estimate, which is based on minimizing the absolute error.

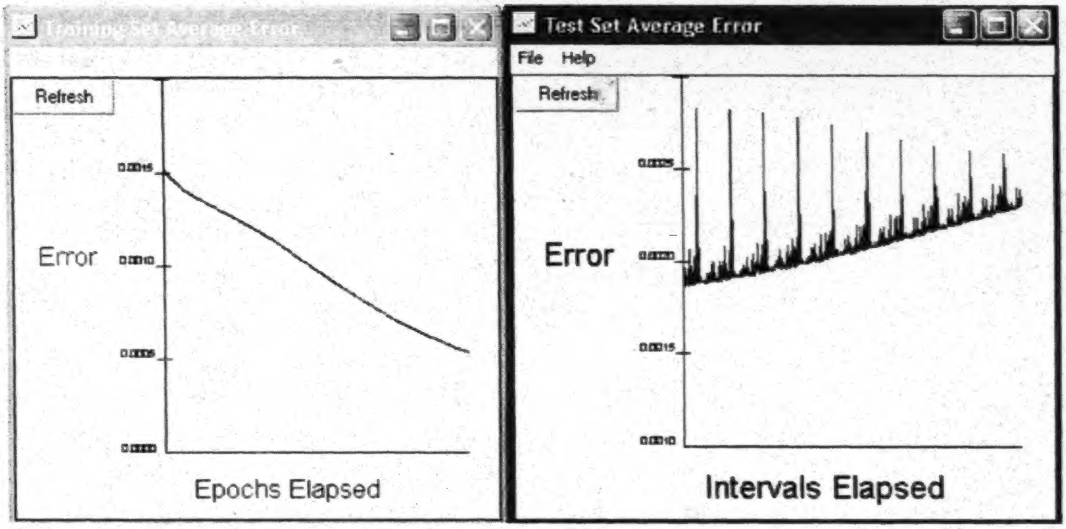
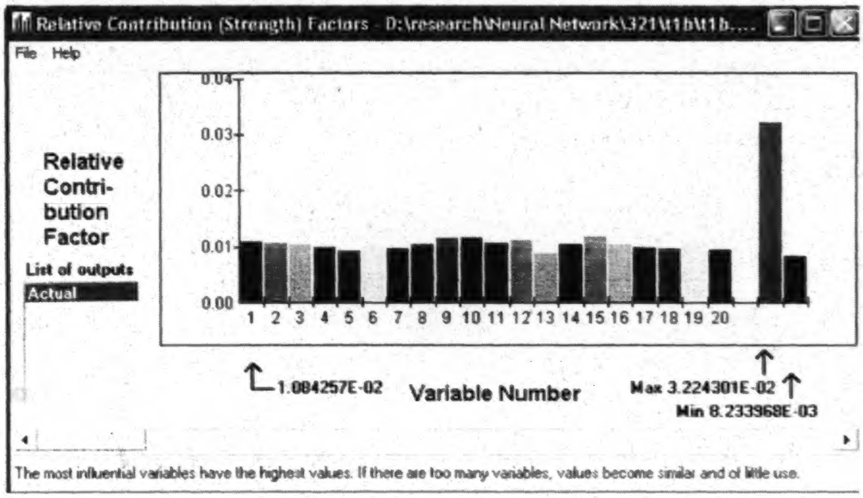

Training Set Average Error
Refresh
Error
0.0015
0.0010
0.0005
0.0000
Epochs Elapsed
Test Set Average Error
File
Help
Refresh
Error
0.0025
0.0020
0.0015
0.0010
Intervals Elapsed

Figure 5.4: Graphical Representation of Training Sets and Intervals Elapsed

Once the learning is complete, the ANN provides a set of contributing factors and their corresponding weights. Figure 5.5 shows a segment of a histogram of the contributing factors based on the input data.


Relative Contribution (Strength) Factors - D:\research\Neural Network\321\t1b\t1b....
File
Help
Relative Contri- bution Factor
List of outputs
Actual
0.04
0.03
0.02
0.01
0.00
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
1.084257E-02
Variable Number
Max 3.224301E-02
Min 8.233968E-03

The most influential variables have the highest values. If there are too many variables, values become similar and of little use.

Figure 5.5: Contributing Factors and Their Weights

Chapter 6 deals with the validation and quality assurance of the contributing factors relative to the trial and error selection of network architectures.

5.6 Output Utilization

The final step is to utilize the data that has been generated through the network model and to determine the confidence level in various classes. The results of the tests that have been performed, using different network architectures, have been tabulated and presented in chapter 6. The best network is selected based on the consistent R^2 value and the minimum average error observed. Neuroshell® 2 ver 4.0 has the ability to generate the source code for the developed network in either of the formats; C Source code (.C) or Visual Basic Source code (.vb) or just extract the formulas for a calculator (.FLA). This entails a much easier pattern in linking the ANN model with the SmartEST by developing a user-friendly interface.

Chapter 6: Observations

6.1 Observations

This chapter deals with the results obtained from the different networks, which have been tabulated with various patterns and worked on while building the ANN model. There were six main network architectures that were tested. The different network architectures are illustrated in Figure 6.1 through Figure 6.6 (*NeuroShell 2, © Ward Systems Group, Inc.)*.

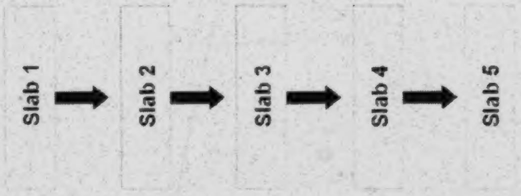

Slab 1
Slab 2
Slab 3
Slab 4
Slab 5

Figure 6.1: Five Layer Standard Connection

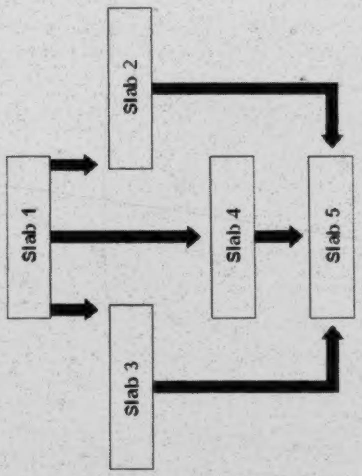

Slab 2
Slab 1
Slab 4
Slab 5
Slab 3

Figure 6.2: Three Hidden Slabs (Ward Nets)

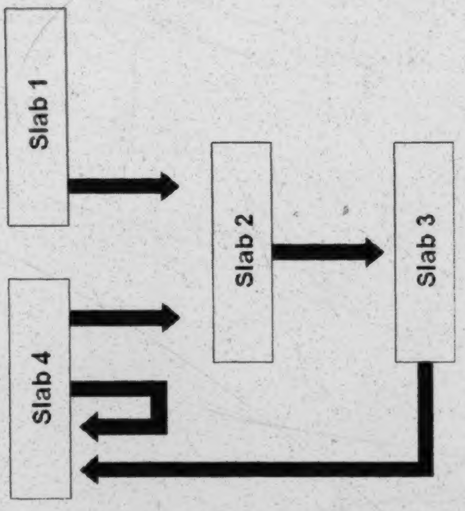

Slab 1
Slab 2
Slab 3
Slab 4

Figure 6.3: Output Layer Dampened Feedback

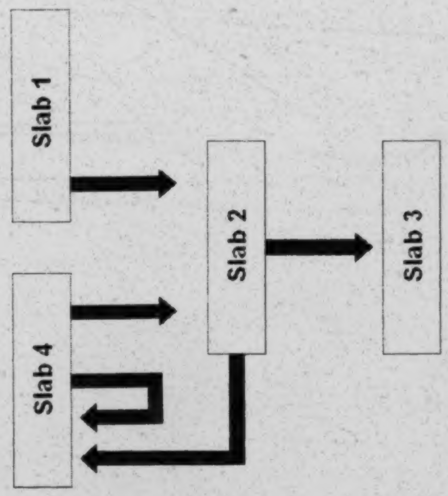

Slab 1
Slab 2
Slab 3
Slab 4

Figure 6.4: Hidden Layer Dampened Feedback

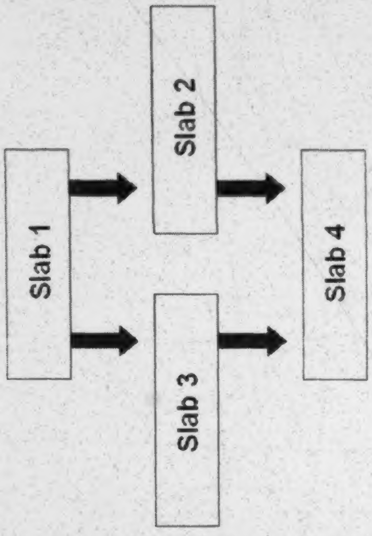

Slab 1
Slab 2
Slab 3
Slab 4

Figure 6.5: Two Hidden Slabs with Different Activation Functions

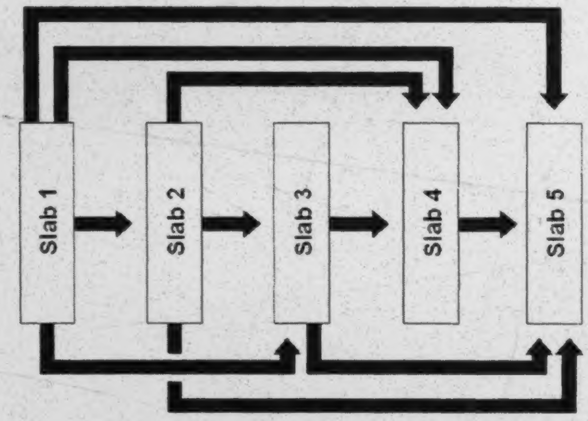

Slab 1
Slab 2
Slab 3
Slab 4
Slab 5

Figure 6.6: Five Layer, Jump Connection

Table 6.1 shows the results obtained from the various network architectures in the 3:1:1 pattern and the best networks have been selected to undergo further scrutiny. The most optimal network is selected on the basis of the best R^2 value of the validation set being compliant with the testing and training sets.

Table 6.1: Network Results for 3:1:1 Pattern

Network Architecture	**Class Pattern**	**Hidden Layer**	**Output Layer**	**Mean Absolute Error($)**	**Training R^2**	**Testing R^2**	**Validation R^2**
Five Layer Standard Connection	3:1:1	Logistic (L.R = 0.1)	Logistic (L.R = 0.1)	4,102.97	0.8235	0.6229	**0.8119**
		Logistic (L.R = 0.2)	Logistic (L.R = 0.2)	4,133.83	0.8238	0.6167	0.8027
		Gaussian (L.R = 0.1)	Gaussian (L.R = 0.1)	4,895.09	0.8235	0.4771	0.575
		Gaussian (L.R = 0.3)	Linear (L.R = 0.3)	8,986.65	-0.0011	0.0001	0.0001
Three Hidden Slab (Ward Nets)	3:1:1	Logistic & Tanh (L.R =0.15)	Logistic (L.R = 0.15) 40,000 int	4,264.92	0.8332	0.5759	0.7801
		Gaussian & Tanh (L.R =0.15)	Logistic (L.R = 0.15) 20,000 int	4,099.75	0.8672	0.5514	0.7461
		Logistic (L.R =0.15)	Logistic (L.R = 0.15) 200,000 int	4,033.34	0.8443	0.6038	**0.8167**
Output Layer Dampened Feedback	3:1:1	**Logistic (L.R = 0.2)**	**Logistic (L.R = 0.2)**	**3,969.80**	**0.829**	**0.625**	**0.8303**
		Gaussian (L.R = 0.2)	Logistic (L.R = 0.2)	4,916.97	0.708	0.5838	0.7285
Hidden Layer Dampened Feedback	3:1:1	Logistic (L.R = 0.2)	Logistic (L.R = 0.2)	4,399.60	0.7193	0.644	**0.7111**
		Gaussian (L.R = 0.2)	Logistic (L.R = 0.2)	5,363.89	0.7908	0.353	0.6199
Two Hidden Slabs , Jump Connection	3:1:1	Logistic & Gaussian (L.R = 0.3)	Logistic (L.R = 0.3) 60,000 int	4,015.27	0.8617	0.5998	0.808
		Gaussian & Gaussian (L.R = 0.3)	Logistic (L.R = 0.3) 200,000 int	4,029.73	0.8651	0.542	0.7861
		Logistic (L.R = 0.3)	Logistic (L.R = 0.3) 60,000 int	4,054.13	0.8206	0.6218	**0.8182**
Five Layers , Jump Connection	3:1:1	Logistic (L.R = 0.3) (Mo = 0.2)	Logistic (L.R = 0.3) 200,000 int	4,181.55	0.7996	0.6091	**0.805**
		Gaussian & logistic (L.R = 0.3) (Mo = 0.2)	Logistic (L.R = 0.3) 200,000 int	4,057.90	0.89	0.5011	0.7622

		Gaussian & tanh (L.R = 0.3) (Mo = 0.2)	Logistic (L.R = 0.3) 200,000 int	4,507.98	0.8139	0.5197	0.7512

In both the 3:1:1 and 3:2:1 patterns of the network architecture that is being designed, the input activation functions are linear satisfying equation 4.

$$\mathbf{F(x) = x} \qquad \mathbf{(4)}$$

$$\text{where } x \leq 0 = 0 \text{ and } 0 < x = 1$$

In the 3:1:1 pattern, the network architecture providing the best optimal solution is the ***"Output layer dampened feedback"*** architecture. These recurrent networks are trained in the same manner as any standard back propagation networks with the exception that the patterns must be presented in the same order i.e., the random selection of pattern is avoided. The difference in the architecture of this network, in comparison to other networks, is the existence of an additional slab (layer) in the input layer that is connected to the hidden layer. This is similar to the initial input slab (layer). This additional slab (layer) holds the contents of a single layer as it existed when the previous pattern was trained. In this way the network sees existing knowledge it had from initial inputs. This slab in the recurrent network architecture serves as the network's "long term" memory *(NeuroShell® 2 Help, ©) Ward Systems Group, Inc.).*

Table 6.2 shows the results obtained from the network architectures utilizing the 3:2:1 pattern as described in chapter 5.

Table 6.2: Network Results for 3:2:1 Pattern

Network Architecture	**Class Pattern**	**Hidden Layer**	**Output Layer**	**Mean Absolute Error**	**Training R^2**	**Testing R^2**	**Validation R^2**
Five Layer Standard Connection	3:2:1	Logistic (L.R = 0.1)	Logistic (L.R = 0.1)	3,809.36	0.8404	0.7163	**0.7758**
		Logistic (L.R = 0.2)	Logistic (L.R = 0.2)	3,841.27	0.8379	0.7251	0.7706
		Gaussian (L.R = 0.1)	Gaussian (L.R = 0.1)	4,505.38	0.8812	0.4566	0.6564
		Gaussian (L.R = 0.3)	Linear (L.R = 0.3)	8,731.07	0.0111	0.0292	0.0463
Three Hidden Slab (Ward Nets)	3:2:1	Gaussian & Tanh (L.R =0.2)	Logistic (L.R = 0.2) 300,000 int	3,691.49	0.9241	0.6157	0.7719
		Logistic & Tanh (L.R =0.2)	Logistic (L.R = 0.2) 300,000 int	3,610.87	0.9015	0.6883	0.7866
		Logistic (L.R =0.2)	**Logistic (L.R = 0.2) 300,000 int**	**3,514.14**	**0.8885**	**0.7225**	**0.7973**
Output Layer Dampened Feedback	3:2:1	Logistic (L.R = 0.1)	Logistic (L.R = 0.1)	3,574.50	0.8876	0.6946	**0.7749**
		Gaussian (L.R = 0.2)	Logistic (L.R = 0.2)	3,724.52	0.9579	0.5019	0.6646
Hidden Layer Dampened Feedback	3:2:1	Logistic (L.R = 0.2)	Logistic (L.R = 0.2)	3,635.76	0.8973	0.6565	**0.7142**
		Gaussian (L.R = 0.2)	Logistic (L.R = 0.2)	4,181.85	0.9735	0.4774	0.6209
		Tanh (L.R = 0.2)	Logistic (L.R = 0.2)	4,535.13	0.9357	0.5970	0.5527
Two Hidden Slabs , Jump Connection	3:2:1	Logistic & Gaussian comp. (L.R = 0.3)	Logistic (L.R = 0.3) 300,000 int	3,879.56	0.9259	0.6348	0.6692
		Gaussian & Gaussian (L.R = 0.3)	Logistic (L.R = 0.3) 300,000 int	3,932.83	0.9335	0.5673	0.6729
		Logistic (L.R = 0.3)	Logistic (L.R = 0.3) 300,000 int	3,686.97	0.8692	0.7138	**0.7777**
		Logistic (L.R = 0.3)	Logistic (L.R = 0.3) 300,000 int	3,755.54	0.8670	0.7157	0.7811

Five Layers , Jump Connection	3:2:1	Logistic (L.R = 0.2)	Logistic (L.R = 0.2) 300,000 int	3,660.97	0.8731	0.7171	**0.7822**
		Gaussian & logistic (L.R = 0.3) (Mo = 0.2)	Logistic (L.R = 0.3) 300,000 int	3,875.29	0.8716	0.6450	0.7865
		Gaussian & tanh (L.R = 0.3) (Mo = 0.2)	Logistic (L.R = 0.3) 300,000 int	3,951.92	0.8740	0.6378	0.7515

As observed from table 6.2, the selected network architecture (highlighted) is the ***"Ward Network"*** (three hidden slabs), which shows a predicted accuracy of **79.73%.** This network architecture was designed by Ward System Group to serve as feature detectors through various activation functions incorporated within the hidden layers. Considering the various R^2 values obtained from the training and the testing of the data among the two pattern sets, it seemed logical to select a network architecture where all the R^2 values of the data sets (training, testing and validation) are close to each other. The tables 6.1 and 6.2 highlight the six main network architectures with at least 2 variations in the activation functions for the hidden and output neuron layers resulting in 17 architecture networks for the 3:1:1 classification and 19 for the 3:2:1 classification.

6.2 Selected Pattern Analysis

Table 6.3 depicts the comparative study between the two various patterns and the various networks incorporated. Figure 6.7 through Figure 6.12 depict the graphic representation of each relation with each other and the R^2 value of the predicted value of the validation set.

Table 6.3: Evaluation of Architectures based on Different Patterns, 3:1:1 and 3:2:1

Network Architectures	3:1:1 Pattern		3:2:1 Pattern	
	Mean Error Values ($)	V - R^2	Mean Error Values ($)	V - R^2
Five Layer Standard Connection	4,102.97	81.19%	3,809.36	77.58%
	4,133.83	80.27%	3,841.27	77.06%
	4,895.09	57.50%	4,505.38	65.64%
	8,986.65	0.01%	8,731.07	4.63%
Three Hidden Slab (Ward Nets)	4,264.92	78.01%	3,610.87	78.66%
	4,099.75	74.61%	3,691.49	77.19%
	4,033.34	81.67%	**3,514.14**	**79.73%**
Output Layer Dampened Feedback	3,969.80	83.03%	3,574.50	77.49%
	4,916.97	72.85%	3,724.52	66.46%
Hidden Layer Dampened Feedback	4,399.60	71.11%	3,635.76	71.42%
	5,363.89	61.99%	4,181.85	62.09%
Two Hidden Slabs , Jump Connection	4,015.27	80.80%	3,879.56	66.92%
	4,029.73	78.61%	3,932.83	67.29%
	4,054.13	81.82%	3,686.97	77.77%
Five Layers ,4 Jump Connection	,181.55	80.50%	3,755.54	78.11%
	4,057.90	76.22%	3,875.29	78.65%
	4,507.98	75.12%	3,951.92	75.15%

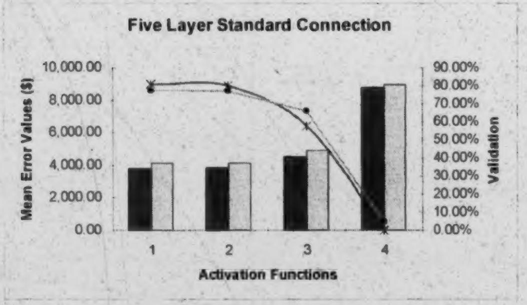


Figure 6.7: Five Layer Standard Connection

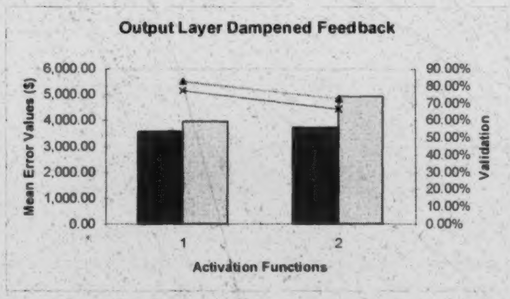


Figure 6.8: Output Layer Dampened Feedback

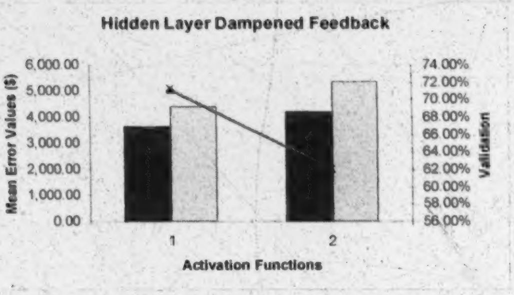


Figure 6.9: Hidden Layer Dampened Feedback

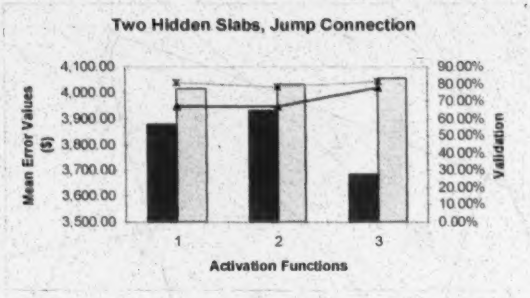


Figure 6.10: Two Hidden Slabs, Different Activation functions

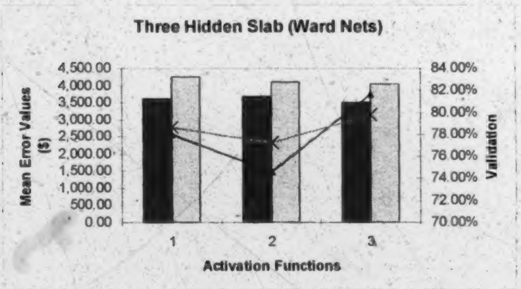


Figure 6.11: Three Hidden Slabs, Ward Nets

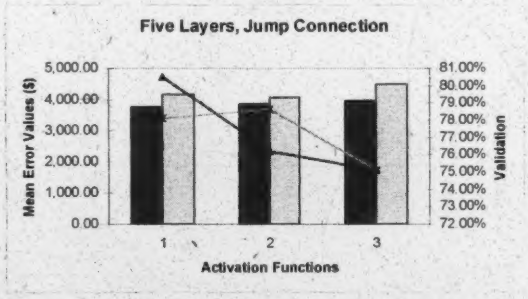


Figure 6.12: Five Layers, Jump Connection

A study of the graphs shown above illustrates that, irrespective of the pattern selected during the development of the Neural Network model, they tend to show similar patterns of behavior within a given network. The main reason of utilizing both patterns is to specifically select the best architecture based on the results incurred. On the basis of the lowest mean error value and the consistency with the R^2 value among the training, testing and validation data, the ***"Ward Network"*** with the three hidden slabs is selected to be best suited for this project. The validation data is then graphed citing out the estimated values with the actual data and the predicted value obtained through the ANN model, which has incorporated the three hidden slab ***"Ward***

Network" architecture. Figure 6.13 shows the graphed illustration of the output for the validation data.

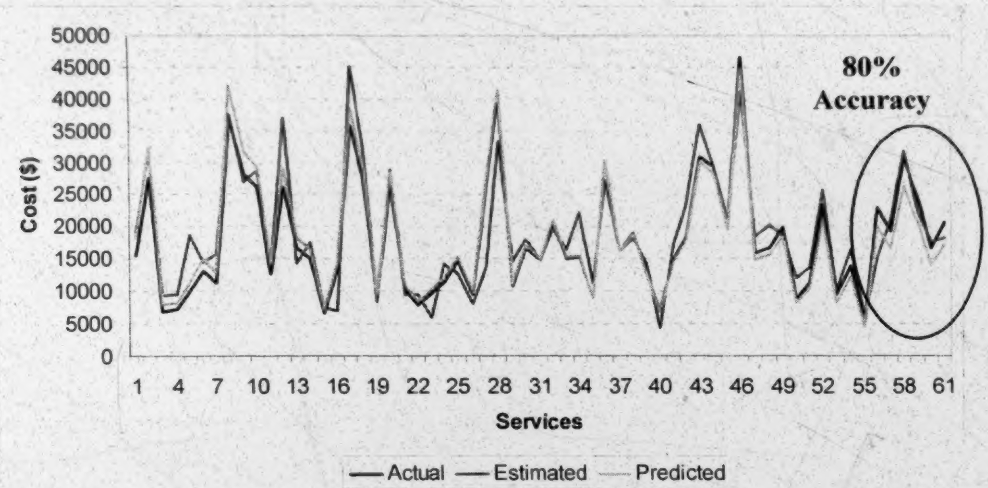


Figure 6.13: Prediction Data Plotted Against the Actual and Estimated Values

It is observed that the pattern predicted shows a close resemblance to the actual cost procured during the duration of the jobs. There are many peaks on the predicted cost value that hit close to the actual values incurred. This is mainly attributed to the fact that the supervised learning approach has been followed with the actual costs incurred serving as the desired output. The highlighted portion depicts the discrepancies that occur in the predicted estimate, which can be attributed to the 20% inaccuracy of the ANN model However, in cases where the variations in cost have been the excess of 60%, there has been a marginal reduction as illustrated in the segment in Figure 6.14 where a previously estimated cost variation of 73.8% has been reduced to 22.1%.

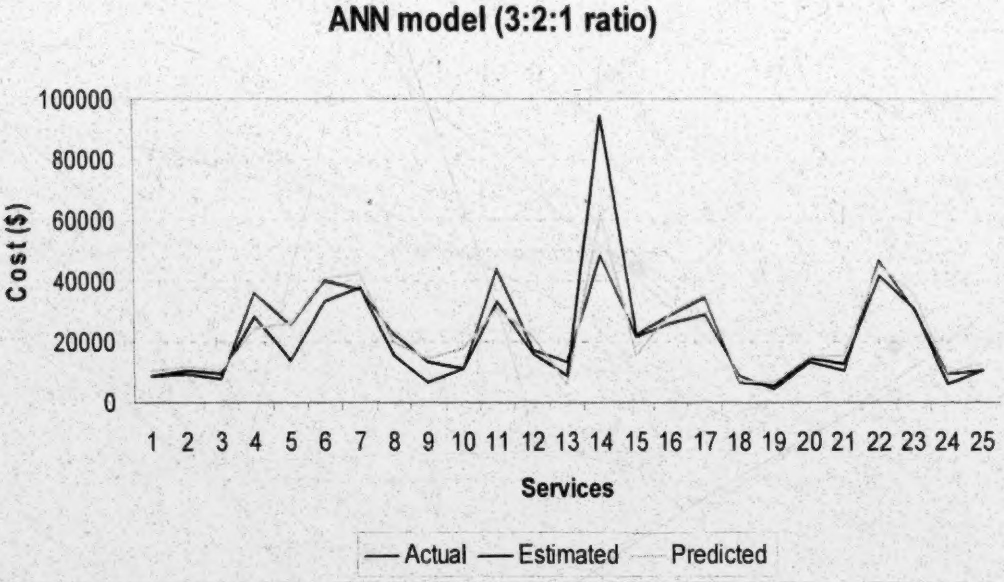


Figure 6.14: Predicted Estimates of Previous Cost Variations over 60%

6.3 Summary and Comments

The justification of the Artificial Neural Network model was done using the validation set of the data obtained from the projects of the City of Edmonton. The aforementioned case study verifies the effectiveness of the input factors that were analyzed and demonstrates the successful implementation of the developed model. The developed ANN model emphasizes the utilization of factors from the environment including the incorporation of a predictive temperature forecasting model in its successful implementation. The advantage of this model is its ability to provide information based on the user's requirement.

Chapter 7: Summary and Conclusions

7.1 Conclusions

Over the last decade, there has been profound utilization of Neural Networks within the construction industry and, in particular, the field of cost estimation. This thesis describes the development of an ANN model for the estimation of costs for water and sewer installation services using the 804 projects obtained from the City of Edmonton. This particular model has obtained all its data from the existing estimating software, the SmartEST, and a predictive temperature forecasting model. The developed ANN model incorporates data relating to the average monthly temperature over the various years, the geographical location of the project, the consistency of the number of estimations performed and features related to site conditions. The ANN model has reduced the variation between the estimated and actual costs and reached a prediction accuracy of about ***80%.***

7.2 Research Contributions

The main contributions of this research can be summarized as follows:

a. The developed Artificial Neural Network model reduces the variation between the estimated and actual costs and reached a prediction accuracy of about 80%

b. The research study shows a correlation between various factors of the environment and the cost estimates, which have been analyzed in detail.

c. The input parameters generated from the research will assist in the development of a user friendly interface to utilize the developed ANN model.

d. The research study also provides an active link in integrating the Artificial Neural Network with the existing estimating software, SmartEST.

e. The study also evaluates the results obtained from the different network architectures by providing a comparative analysis.

The state of the art Artificial Neural Network model designed in this study emphasizes on the use of the estimate values of labor, material, equipment and other costs as input parameters for the development of the neural network model. The predicted estimate thus obtained through the ANN model enhances the accuracy of the estimate done through the SmartEST. The study ascertains the validation of the selection of the network architecture based on the classification of the data sets in the two ratios 3:1:1 and 3:2:1. The results obtained from both classifications have been tabulated in tables 6.1 and 6.2 in chapter 6 and this study was performed to select the best optimum network, not just on the basis of the lowest mean error value and the higher R^2 value of the validation sets but deals with the consistency between the R^2 values of the training and testing data sets with the R^2 values of the validation sets. The validation of the selection of the ***"Ward Network"*** was done on the basis of the comparative study of the different network architectures in both sets of classification of data in the ratios 3:1:1 and 3:2:1 as shown in figures 6.7 through till 6.12 in chapter 6. In addition, the developed Neural Network model is designed to incorporate changes within the network architecture based on the data obtained during real time operation within the field.

7.3 Recommendations for the Artificial Neural Network Model

The City of Edmonton has intentions of incorporating lean thinking within the water and sewer installation services in the near future. The ANN model developed can still help set the baseline for a much more accurate estimate based on modifying the existing network architecture.

The current input factors included dividing the city into four sectors namely: South West, South East, North East and North West. Distribution of the services into further sectors would give a better understanding of how the city relates to location and, in particular, the geological issues arising from the various land patterns. Based on the current information gathered, if the effect of geotechnical land data is available for the various sectors, this would definitely enhance the prediction of the Neural Network.

A user friendly interface would also assist in the prediction of the cost based on projects within or on time, or projects that are delayed over certain time frames. This sort of development could gradually allow the network to field data and incorporate costs accordingly.

7.4 Recommendation for Future Research

Figure 1.2 in chapter 1 illustrates the future work that needs to be done for the full functioning of the network with the existing estimating software, the SmartEST to reduce the percentage of those projects that have been unevenly estimated. The immediate factors that need to be considered include:

a. Obtain the data from 2005 and include them into the network and see how they coexist with the current network.
b. Retrain the network with the new data obtained and discard the older data over a 5 year time period.
c. Real time operation of the data with the next year and train the data according to the information obtained.
d. Based on the improvement of the process productivity through lean principles, make alterations within the network to work accordingly.
e. Build a user friendly interface, preferably a visual basic platform, to link the SmartEST with the Artificial Neural Network model.

References

Rosenblatt,F. (1962). Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms

Tsoulakas and Uhrig (1997). Fuzzy and Neural Approaches in Engineering

Minsky,M. and Papert,S. (1987). Perceptrons - Expanded Edition: An Introduction to Computational Geometry.

Haykin, S. (1994). Neural Networks: A Comprehensive Foundation

McMurrey,D. A.(2001). Power Tools for Technical Communication

Tabachnick and Fidel (1989). Using multivariate statistics. (2nd edition). New York: HarperCollins

Smith, L., University of Stirling (1996). An Introduction to Neural Network, Department of Computing and Mathematics.

Chao, L., and Skibniewski, M. (1994). "Estimating construction productivity: Neural-network-based approach." Journal of Computing in Civil Engineering, ASCE, 8(2), 234–251.

Portas, J., and AbouRizk, S. (1997). "A Neural Network Model for Estimating Construction Productivity" Journal of Construction Engineering and Management, *ASCE, Volume 123, No. 4, pp. 399-410, December 1997.*

AbouRizk, S., Knowles, P., and Hermann, U. (2001). "Estimating Labor Productivity for Industrial Construction Activities." Journal of Construction Engineering and Management, Vol 127, No 6, pp 502 - 511, Nov./Dec. 2001.

Lu, M, AbouRizk, S, Hermann, U (2000) "Estimating Labor Productivity Using Probability Inference Neural Network", Journal of Computing in Civil Engineering, ASCE, Volume 14, Issue 4, pp. 241-248 (October 2000)

Lu, M., AbouRizk, S., and Hermann, U. (2001). "Sensitivity Analysis of Neural Networks in Spool Fabrication Productivity Studies", Journal of Computing in Civil Engineering, ASCE, Volume 15, No. 4, pp. 299-308, October 2001.

Forbes, D.R., McGurnaghan, H., Graham, D., Smith, S.D. (2004); "Concrete Placing Productivity Using a Novel Neural Network Design", ARCOM 20th Annual Conference, Heriot-Watt University, Sept 1-3, 1053-1062.

Rowings, J.E, Sonmez, R (1998); "Construction Labor Productivity Modeling with Neural Networks", Journal of Construction Engineering and Management. Volume 124, Issue 6, pp. 498-504

Koskela, L. (2000). An exploration towards a production theory and its application to construction; VVT Technical Research Center of Finland

Duff, R., Emsley, M., Gregory, M., Lowe, D. and Masterman, J. (1998). Development of a model of total building procurement costs for construction clients; 14th annual conference, ARCOM, Reading, 210-218

Harding, A., Lowe, D., Emsley, M., Hickson, A. and Duff,R., Department of Building Engineering, UMIST, UK (1999). The Role of Neural networks in early stage Cost Estimation in the 21st Century.

Smith, A.E and Mason, A.K (1996) "Cost Estimation Predictive Modeling: Regression versus Neural Network," accepted to The Engineering Economist, November edition, vol 42, 1996.

Brass J., Gerrard, A.M. and Peel, D. (1994) "Estimating vessel costs via neural networks," Proceedings of the 13th International Cost Engineering Congress, London.

Gerrard, A.M., Brass, J. and Peel, D. (1994) "Using neural nets to cost chemical plants," Proceedings of the 4th European Symposium on Computer-Aided Process Engineering, 475-478.

de la Garza, J. M. and Rouhana, K. G. (1995) "Neural networks versus parameter-based applications in cost estimating," Cost Engineering, vol. 37, no. 2, 14-18.

Shtub, A. and Zimmerman, Y. (1993) "Neural-network-based approach for estimating the cost of assembly systems," International Journal of Production Economics, vol. 32.

El- Gafy, M.A.H, Ibrahim, M.E., Taha, M.A (2001) "Neural Network Model for Parametric Cost Estimation of Sewer Projects", Master of Sciences Thesis Report, Cairo University, Egypt

Siqueira, I, Moselhi, O (1999) "Neural Network based Cost estimating" , Master of Applied Sciences Thesis Report, Concordia University, Montreal, Canada

Sodikov, J., Department of Civil and Environmental Eng., Saitama University , Japan

(2005).Cost Estimation of Highway Projects in Developing Countries: Artificial Neural Network Approach; Journal of the Eastern Asia Society for Transportation Studies, Vol. 6, pp. 1036 - 1047, 2005

Schexnayder, C. J. and Mayo, R. E., (2003), Construction Management Fundamentals, McGraw-Hill Higher Education, Boston, MA

Kung, D. (2005). Process and productivity improvement study for the sewer and water service installation by the City of Edmonton construction crews; M.Eng report, University of Alberta.

[Web 1] *Chicurel, M. (1995). The Inner Life of Neurons, <http://www.med.harvard.edu/publications/On_The_Brain/Volume4/Number2/SP95In.html > last visited on December 1st 2005.*

[Web 2] *Andras ,P. (1997). Artificial Neural Networks – Introduction, <http://www.staff.ncl.ac.uk/peter.andras/annintro.ppt> last visited on Dec 1st 2005*

[Web 3] *Stergiou,C. and Siganos.D; Imperial College, UK. Neural Networks,*

<http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html> last visited on December 1st 2005

[Web 4] *Rodríguez, F.J.; University of Strahclyde); Dr. Sandham,W.A.; ERASMUS program, European Union. Artificial Neural Networks Tutorials, <http://www.gc.ssr.upm.es/inves/neural/ann1/concepts/Suunsupm.htm> last visited on December 1st, 2005*

[Web 5] *Habra, A. (2005). Neural Networks – An Introduction, Abdul Habra; <http://www.tek271.com/articles/neuralNet/IntoToNeuralNets.html > last visited on December 1st 2005*

[Web 6] *Yilmas, C. (2003). The International Journal of Artificial Organs / Vol. 28 / no. 1, 2005 / pp. 1-2, <http://www.artificialorgans.com/content/210/ 2180/7511.pdf> last visited on December 1st 2005*

[Web 7] *Courtesy: Environment Canada: Edmonton Temperature Data 2005 <http://www.climate.weatheroffice.ec.gc.ca/climateData> last visited on March 6th 2005.*

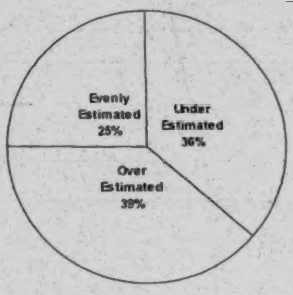
[Web 8] *Courtesy: www.mapquest.ca <http://www.mapquest.com/maps/map.adp?formtype=address&country=CA&popflag=0&latitude=&longitude=&name=&phone=&level=&addtohistory=&cat=&address=&city=Edmonton&state=Ab&zipcode=> last visited on December 1st 2005.*

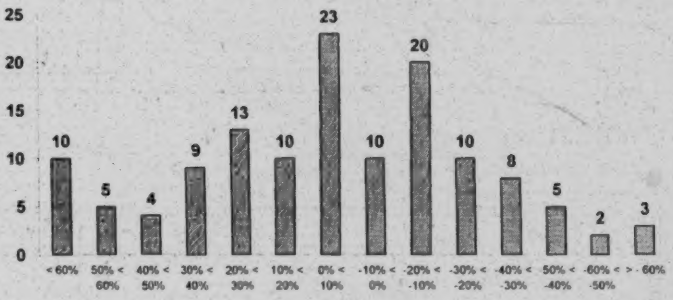
Neuroshell® 2 release 4.0 Copyright© 1993-2000, Ward Systems Group® Inc. <http://www.wardsystems.com>

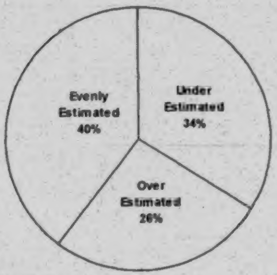
El-Kassas, E. M. A., Mackie, R. I. and El-Sheikh, A. I., Using Neural Networks in Cold-Formed Steel Design; Submitted for publication in the International Journal of Computers and Structures

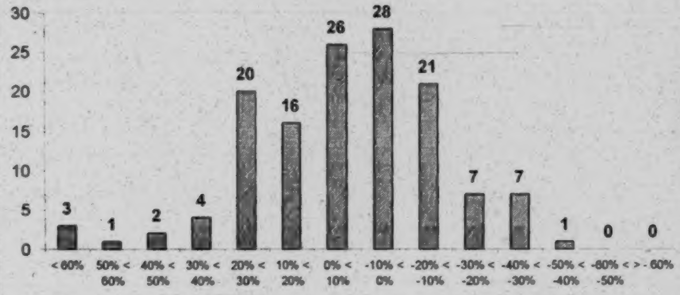
Li, Tie; Ma, Guang; Ying, Fang; Yu, Haitao (2005). Temperature Forecasting Model, an unpublished manuscript submitted in class.

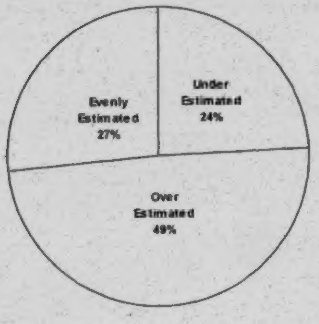
Appendix A

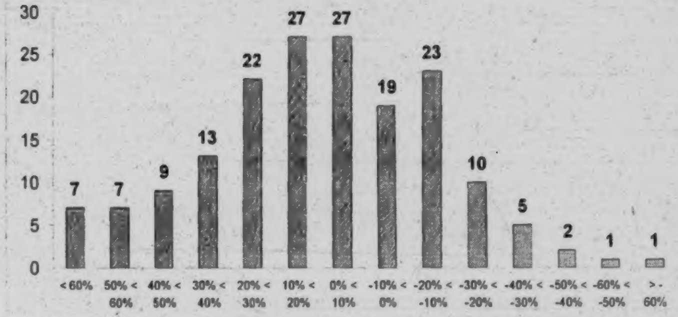

2004 Estimated Costs for Service Installations

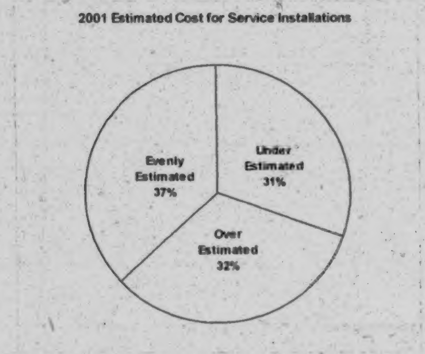

2004 Service Installations

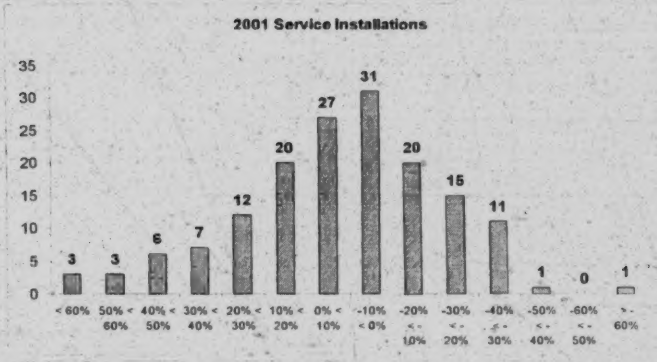

2003 Estimated Cost for Service Installations

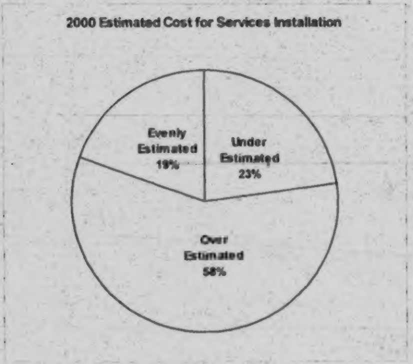

2003 Services Installation

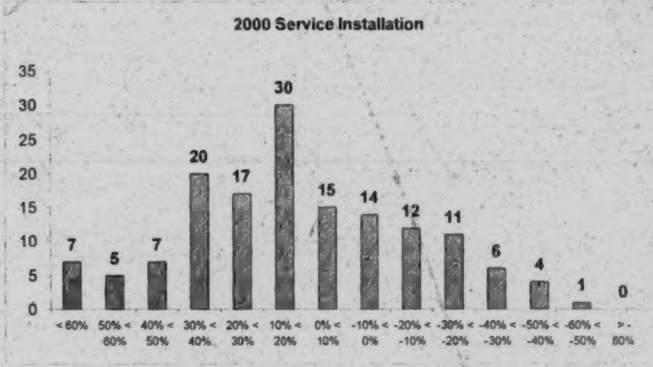

2002 Estimated Cost for Service Installations

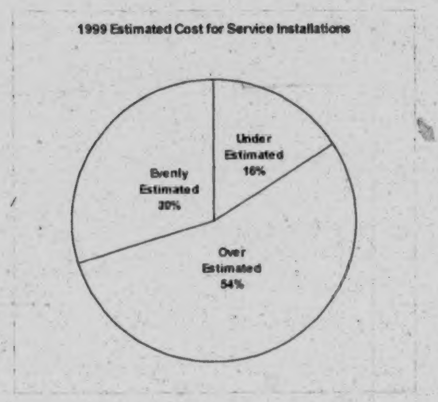

2002 Service Installations

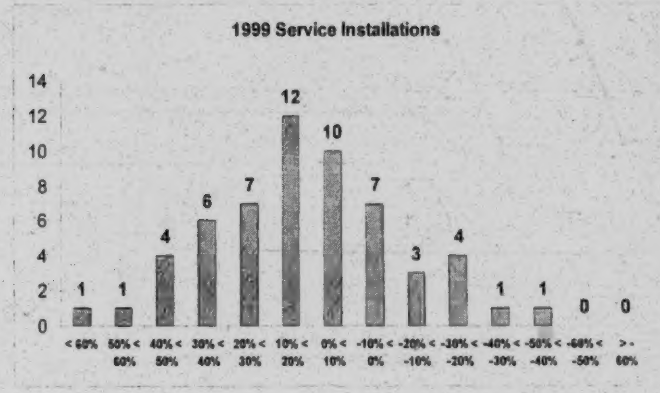

2001 Estimated Cost for Service Installations
Evenly Estimated 37%
Under Estimated 31%
Over Estimated 32%


2001 Service Installations
35
30
25
20
15
10
5
0
3
3
6
7
12
20
27
31
20
15
11
1
0
1
< 60%
50% < 60%
40% < 50%
30% < 40%
20% < 30%
10% < 20%
0% < 10%
-10% < 0%
-20% < -10%
-30% < -20%
-40% < -30%
-50% < -40%
-60% < -50%
> -60%

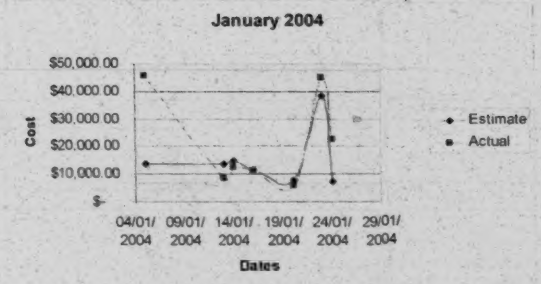

2000 Estimated Cost for Services Installation
Evenly Estimated 19%
Under Estimated 23%
Over Estimated 58%

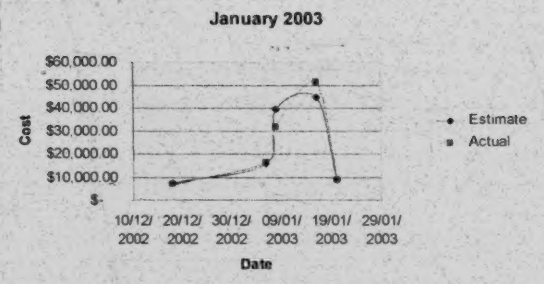

2000 Service Installation
35
30
25
20
15
10
5
0
7
5
7
20
17
30
15
14
12
11
6
4
1
0
< 60%
50% < 60%
40% < 50%
30% < 40%
20% < 30%
10% < 20%
0% < 10%
-10% < 0%
-20% < -10%
-30% < -20%
-40% < -30%
-50% < -40%
-60% < -50%
> -60%

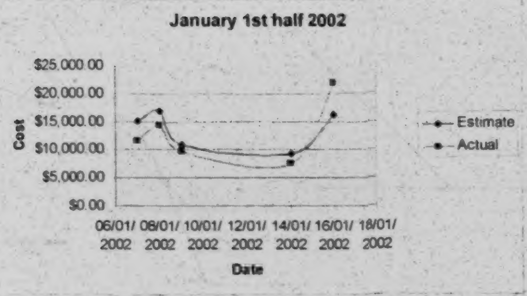

1999 Estimated Cost for Service Installations
Evenly Estimated 30%
Under Estimated 16%
Over Estimated 54%

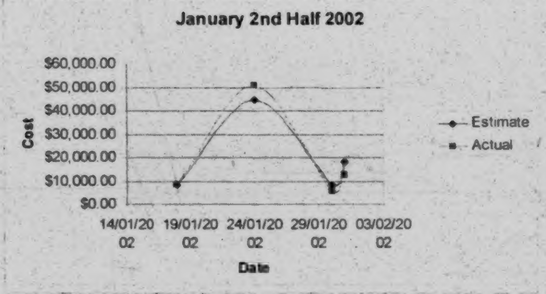

1999 Service Installations
14
12
10
8
6
4
2
0
1
1
4
6
7
12
10
7
3
4
1
1
0
0
< 60%
50% < 60%
40% < 50%
30% < 40%
20% < 30%
10% < 20%
0% < 10%
-10% < 0%
-20% < -10%
-30% < -20%
-40% < -30%
-50% < -40%
-60% < -50%
> -60%

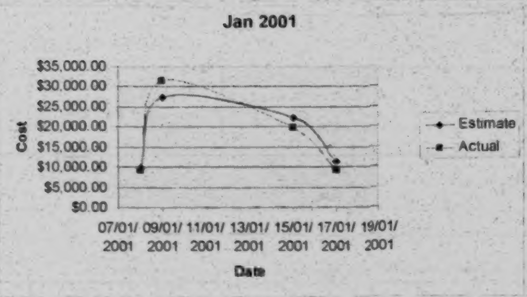
Appendix B

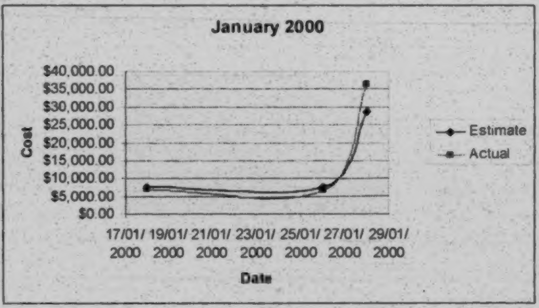

January 2004
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$-
Cost
04/01/ 2004
09/01/ 2004
14/01/ 2004
19/01/ 2004
24/01/ 2004
29/01/ 2004
Dates
Estimate
Actual

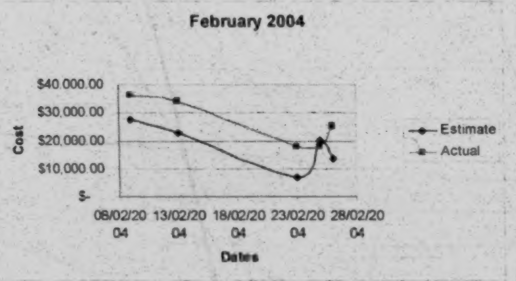

January 2003
$60,000.00
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$-
Cost
10/12/ 2002
20/12/ 2002
30/12/ 2002
09/01/ 2003
19/01/ 2003
29/01/ 2003
Date
Estimate
Actual

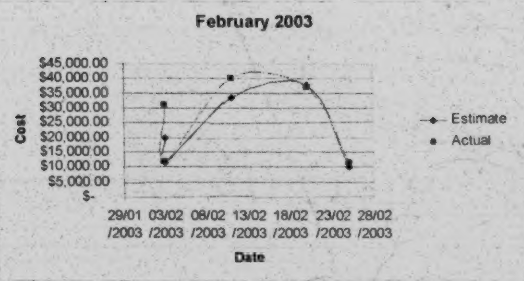

January 1st half 2002
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
Cost
06/01/ 2002
08/01/ 2002
10/01/ 2002
12/01/ 2002
14/01/ 2002
16/01/ 2002
18/01/ 2002
Date
Estimate
Actual

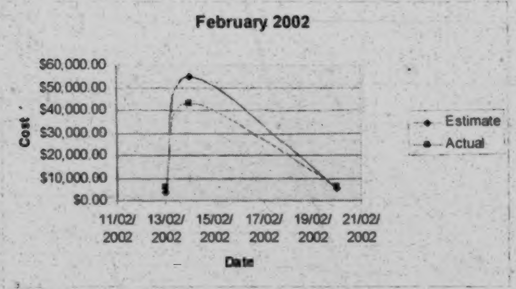

January 2nd Half 2002
$60,000.00
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$0.00
Cost
14/01/20 02
19/01/20 02
24/01/20 02
29/01/20 02
03/02/20 02
Date
Estimate
Actual

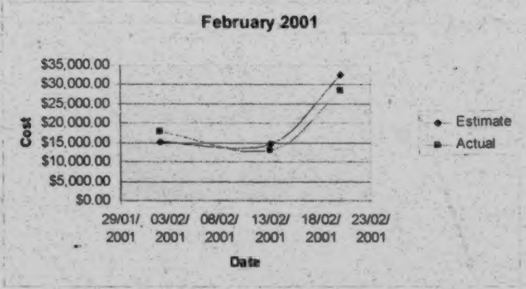

Jan 2001
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
Cost
07/01/ 2001
09/01/ 2001
11/01/ 2001
13/01/ 2001
15/01/ 2001
17/01/ 2001
19/01/ 2001
Date
Estimate
Actual

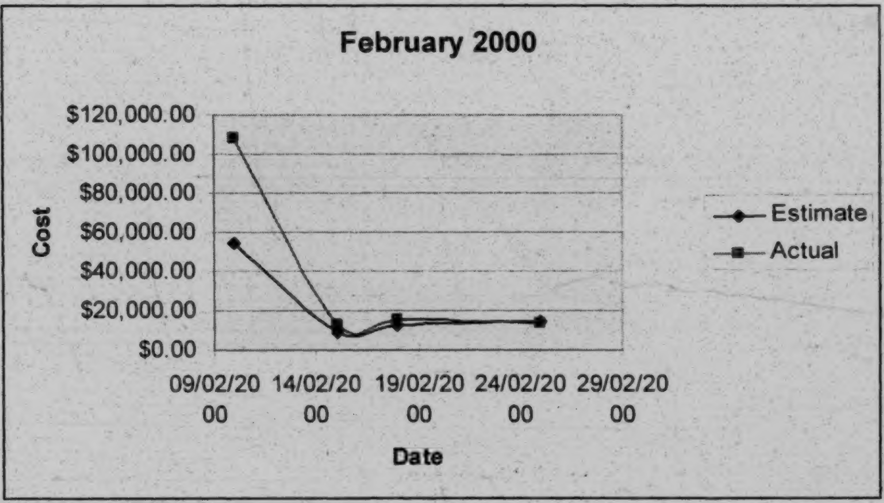

January 2000
$40,000.00
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
Cost
17/01/ 2000
19/01/ 2000
21/01/ 2000
23/01/ 2000
25/01/ 2000
27/01/ 2000
29/01/ 2000
Date
Estimate
Actual

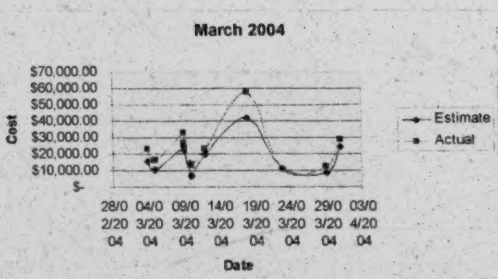

February 2004
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$-
Cost
08/02/20 04
13/02/20 04
18/02/20 04
23/02/20 04
28/02/20 04
Dates
Estimate
Actual

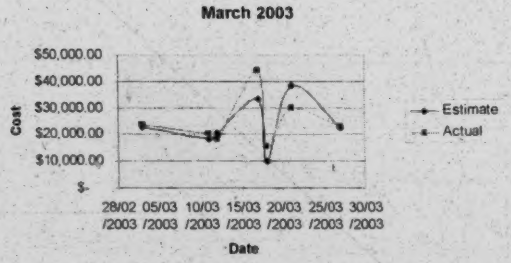

February 2003
$45,000.00
$40,000.00
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$-
Cost
29/01 /2003
03/02 /2003
08/02 /2003
13/02 /2003
18/02 /2003
23/02 /2003
28/02 /2003
Date
Estimate
Actual

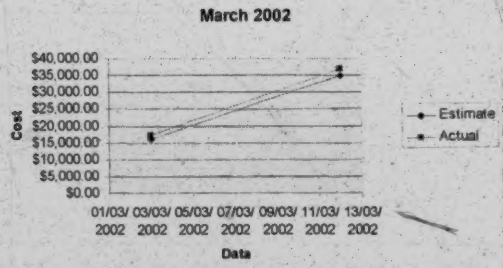

February 2002
$60,000.00
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$0.00
Cost
11/02/ 2002
13/02/ 2002
15/02/ 2002
17/02/ 2002
19/02/ 2002
21/02/ 2002
Date
Estimate
Actual

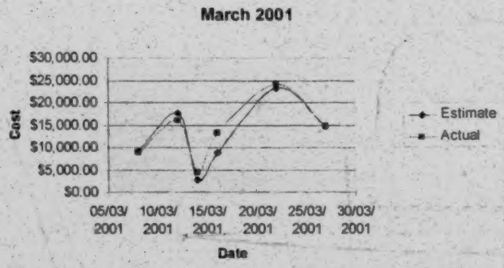

February 2001
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
Cost
29/01/ 2001
03/02/ 2001
08/02/ 2001
13/02/ 2001
18/02/ 2001
23/02/ 2001
Date
Estimate
Actual

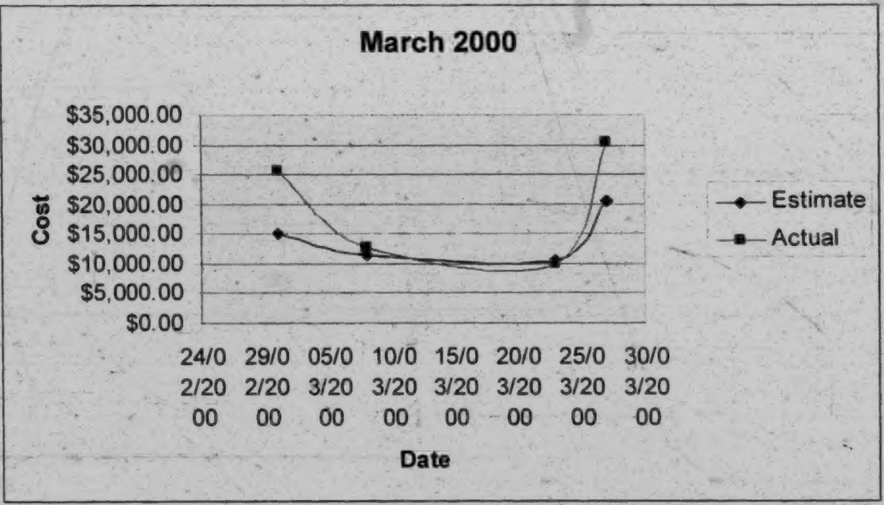

February 2000
$120,000.00
$100,000.00
$80,000.00
$60,000.00
$40,000.00
$20,000.00
$0.00
Cost
09/02/20 00
14/02/20 00
19/02/20 00
24/02/20 00
29/02/20 00
Date
Estimate
Actual

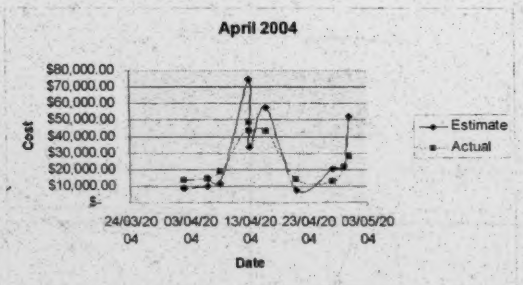

March 2004
$70,000.00
$60,000.00
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$-
Cost
28/02/2004
04/03/2004
09/03/2004
14/03/2004
19/03/2004
24/03/2004
29/03/2004
03/04/2004
Date
Estimate
Actual

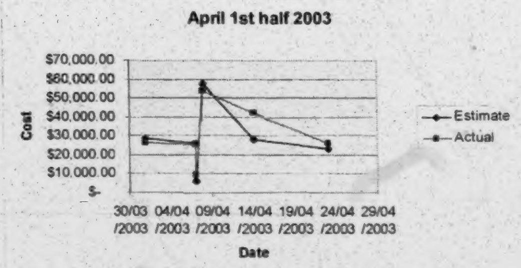

March 2003
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$-
Cost
28/02/2003
05/03/2003
10/03/2003
15/03/2003
20/03/2003
25/03/2003
30/03/2003
Date
Estimate
Actual

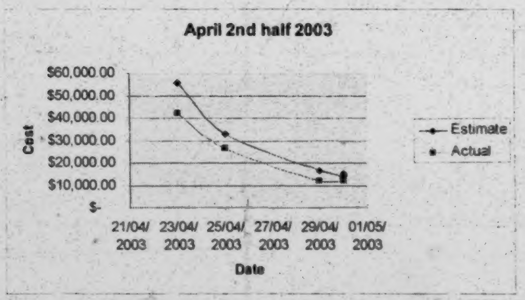

March 2002
$40,000.00
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
Cost
01/03/2002
03/03/2002
05/03/2002
07/03/2002
09/03/2002
11/03/2002
13/03/2002
Data
Estimate
Actual

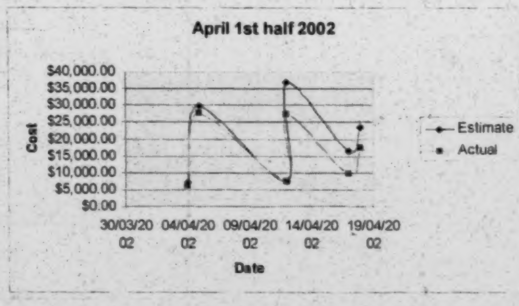

March 2001
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
Cost
05/03/2001
10/03/2001
15/03/2001
20/03/2001
25/03/2001
30/03/2001
Date
Estimate
Actual

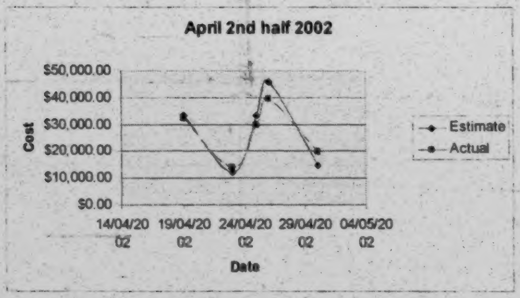

March 2000
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
Cost
24/02/2000
29/02/2000
05/03/2000
10/03/2000
15/03/2000
20/03/2000
25/03/2000
30/03/2000
Date
Estimate
Actual

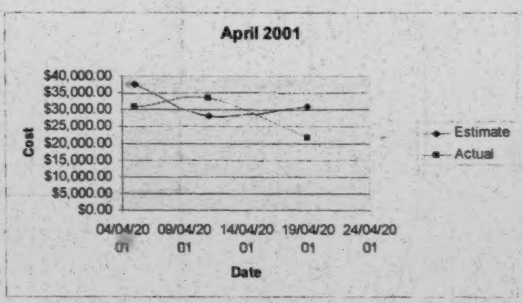

April 2004
Cost
$80,000.00
$70,000.00
$60,000.00
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$-
24/03/2004
03/04/2004
13/04/2004
23/04/2004
03/05/2004
Date
Estimate
Actual

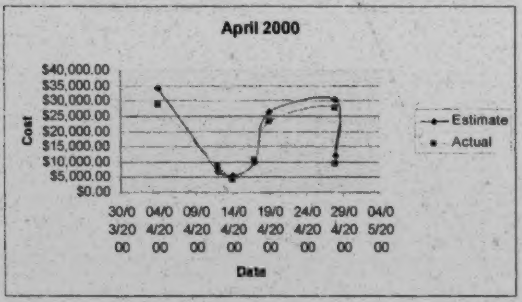

April 1st half 2003
Cost
$70,000.00
$60,000.00
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$-
30/03/2003
04/04/2003
09/04/2003
14/04/2003
19/04/2003
24/04/2003
29/04/2003
Date
Estimate
Actual

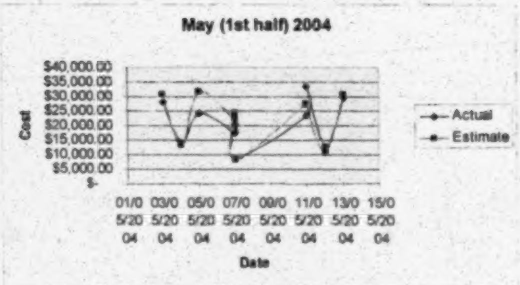

April 2nd half 2003
Cost
$60,000.00
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$-
21/04/2003
23/04/2003
25/04/2003
27/04/2003
29/04/2003
01/05/2003
Date
Estimate
Actual

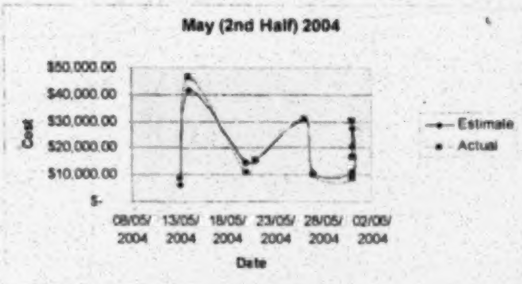

April 1st half 2002
Cost
$40,000.00
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
30/03/2002
04/04/2002
09/04/2002
14/04/2002
19/04/2002
Date
Estimate
Actual

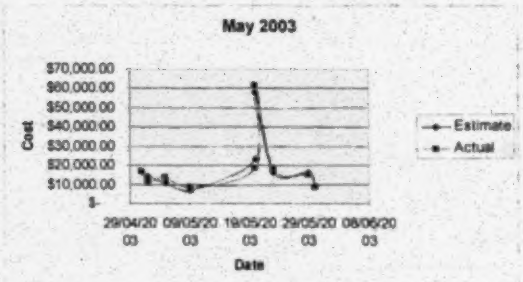

April 2nd half 2002
Cost
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$0.00
14/04/2002
19/04/2002
24/04/2002
29/04/2002
04/05/2002
Date
Estimate
Actual

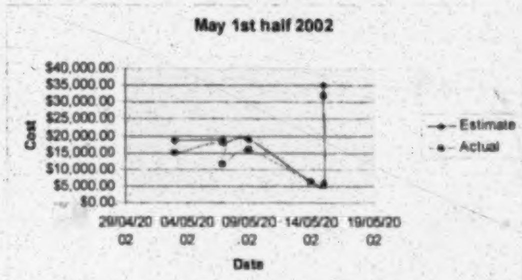

April 2001
Cost
$40,000.00
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
04/04/2001
09/04/2001
14/04/2001
19/04/2001
24/04/2001
Date
Estimate
Actual

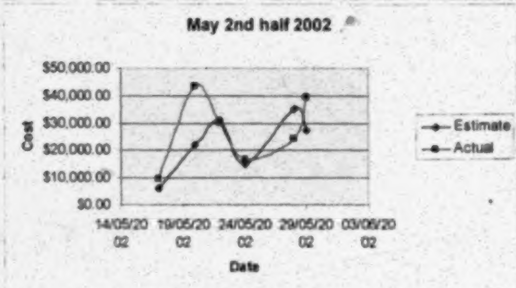

April 2000
Cost
$40,000.00
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
30/03/2000
04/04/2000
09/04/2000
14/04/2000
19/04/2000
24/04/2000
29/04/2000
04/05/2000
Date
Estimate
Actual

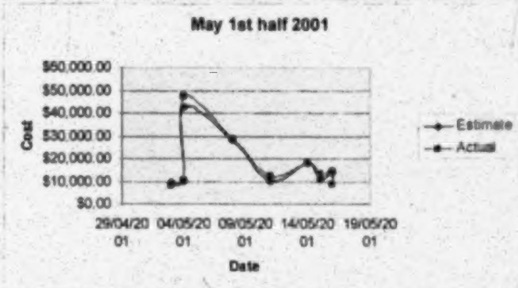

May (1st half) 2004

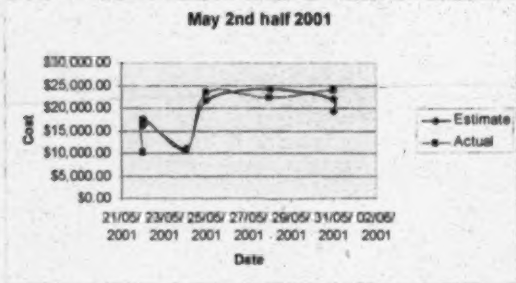

May (2nd Half) 2004

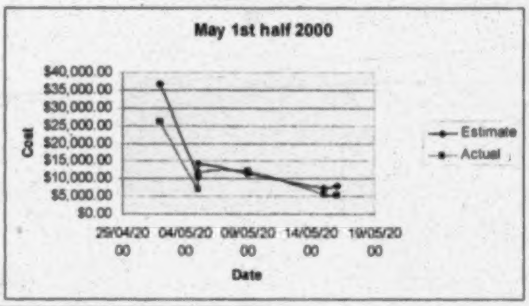

May 2003

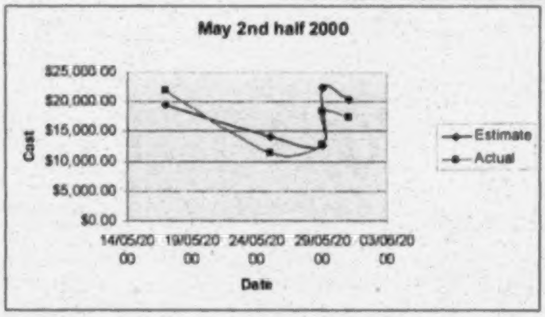

May 1st half 2002

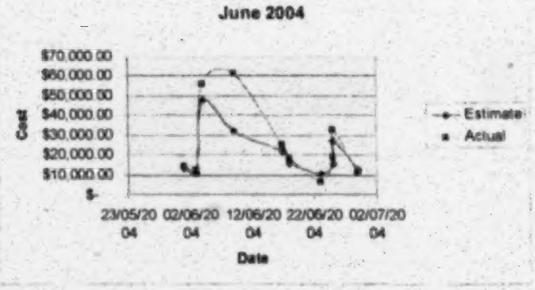

May 2nd half 2002

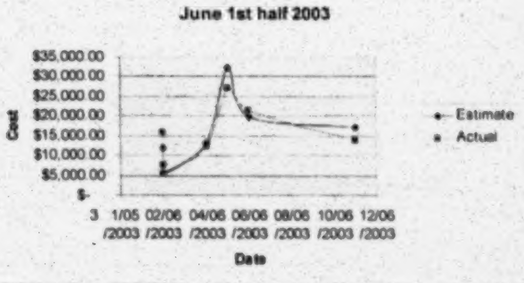

May 1st half 2001

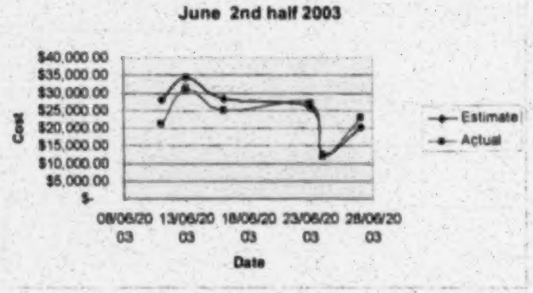

May 2nd half 2001

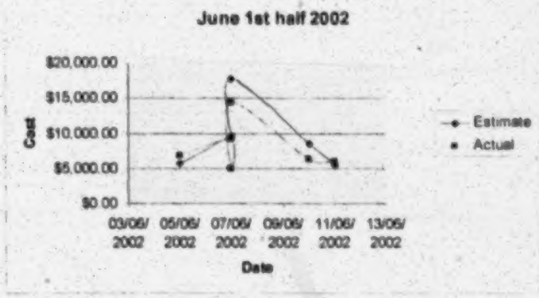

May 1st half 2000

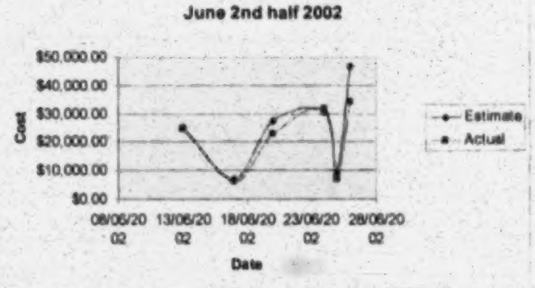

May 2nd half 2000

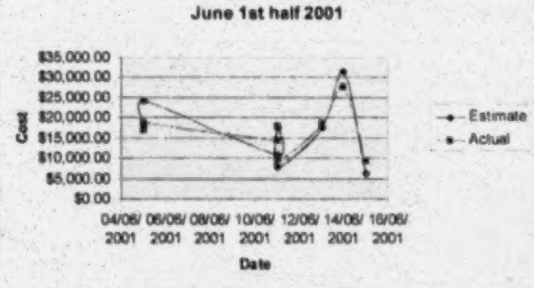

June 2004
$70,000.00
$60,000.00
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$-
Cost
23/05/2004
02/06/2004
12/06/2004
22/06/2004
02/07/2004
Date
Estimate
Actual

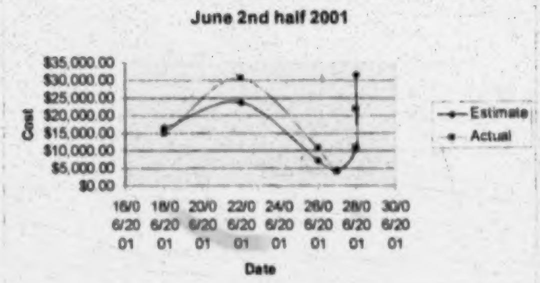

June 1st half 2003
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$-
Cost
31/05/2003
02/06/2003
04/06/2003
06/06/2003
08/06/2003
10/06/2003
12/06/2003
Date
Estimate
Actual

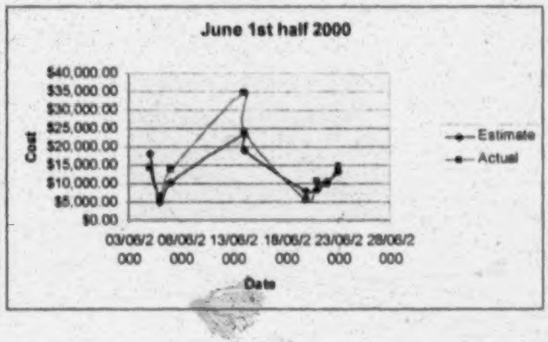

June 2nd half 2003
$40,000.00
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$-
Cost
08/06/2003
13/06/2003
18/06/2003
23/06/2003
28/06/2003
Date
Estimate
Actual

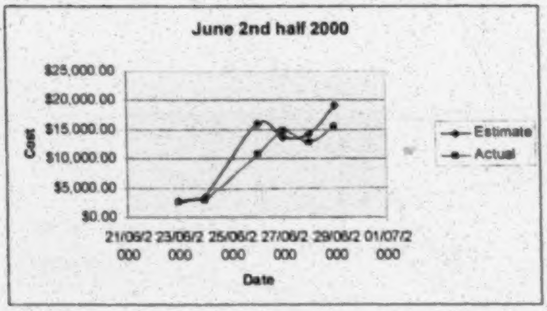

June 1st half 2002
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
Cost
03/06/2002
05/06/2002
07/06/2002
09/06/2002
11/06/2002
13/06/2002
Date
Estimate
Actual

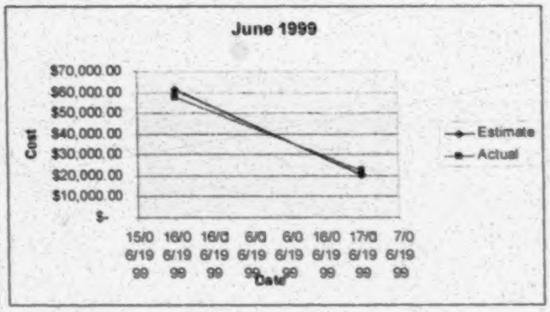

June 2nd half 2002
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$0.00
Cost
08/06/2002
13/06/2002
18/06/2002
23/06/2002
28/06/2002
Date
Estimate
Actual

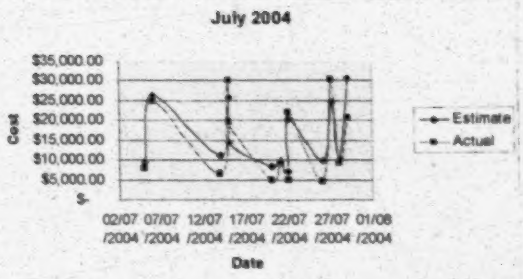

June 1st half 2001
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
Cost
04/06/2001
06/06/2001
08/06/2001
10/06/2001
12/06/2001
14/06/2001
16/06/2001
Date
Estimate
Actual

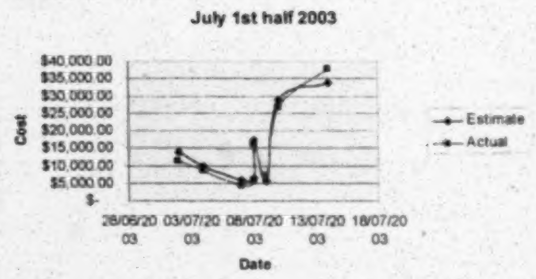

June 2nd half 2001
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
Cost
16/06/2001
18/06/2001
20/06/2001
22/06/2001
24/06/2001
26/06/2001
28/06/2001
30/06/2001
Date
Estimate
Actual

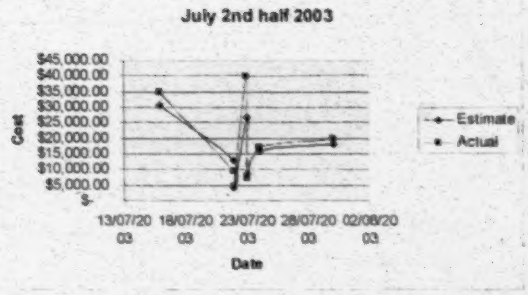

June 1st half 2000
$40,000.00
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
Cost
03/06/2000
08/06/2000
13/06/2000
18/06/2000
23/06/2000
28/06/2000
Date
Estimate
Actual

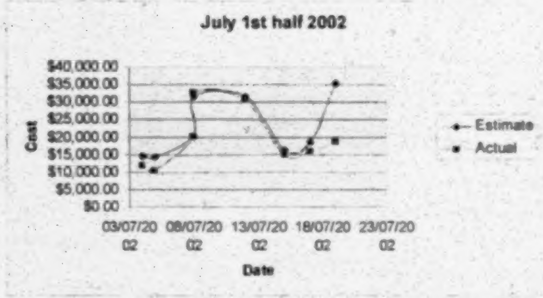

June 2nd half 2000
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
Cost
21/06/2000
23/06/2000
25/06/2000
27/06/2000
29/06/2000
01/07/2000
Date
Estimate
Actual

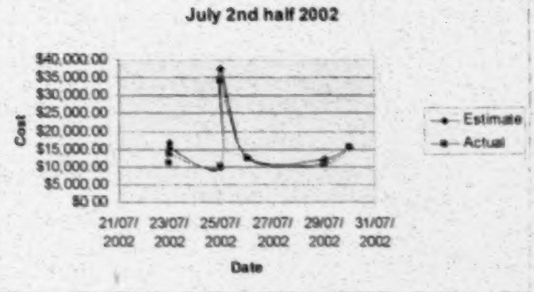

June 1999
$70,000.00
$60,000.00
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$-
Cost
15/06/1999
16/06/1999
16/06/1999
6/06/1999
6/06/1999
16/06/1999
17/06/1999
7/06/1999
Date
Estimate
Actual

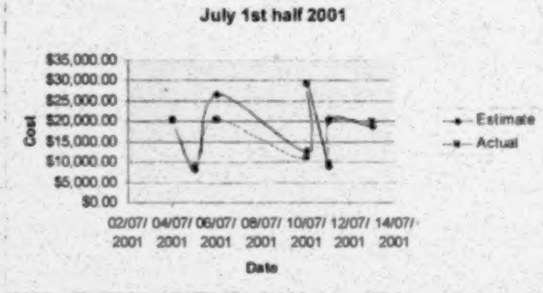

July 2004
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$-
Cost
02/07 /2004
07/07 /2004
12/07 /2004
17/07 /2004
22/07 /2004
27/07 /2004
01/08 /2004
Date
Estimate
Actual

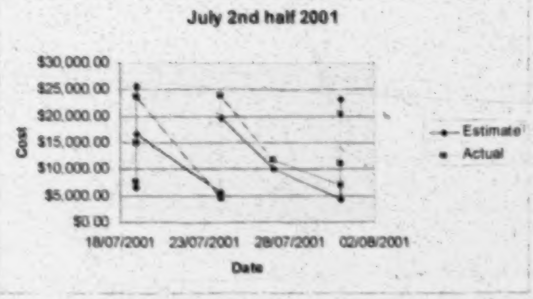

July 1st half 2003
$40,000.00
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$-
Cost
28/06/20 03
03/07/20 03
08/07/20 03
13/07/20 03
18/07/20 03
Date
Estimate
Actual

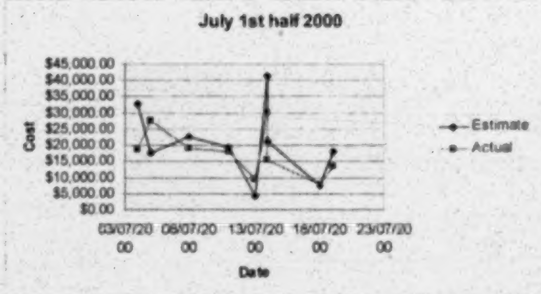

July 2nd half 2003
$45,000.00
$40,000.00
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$-
Cost
13/07/20 03
18/07/20 03
23/07/20 03
28/07/20 03
02/08/20 03
Date
Estimate
Actual

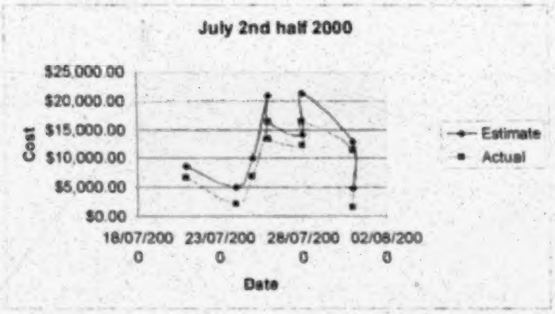

July 1st half 2002
$40,000.00
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
Cost
03/07/20 02
08/07/20 02
13/07/20 02
18/07/20 02
23/07/20 02
Date
Estimate
Actual

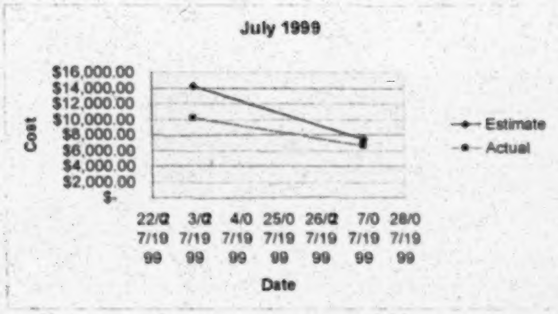

July 2nd half 2002
$40,000.00
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
Cost
21/07/ 2002
23/07/ 2002
25/07/ 2002
27/07/ 2002
29/07/ 2002
31/07/ 2002
Date
Estimate
Actual

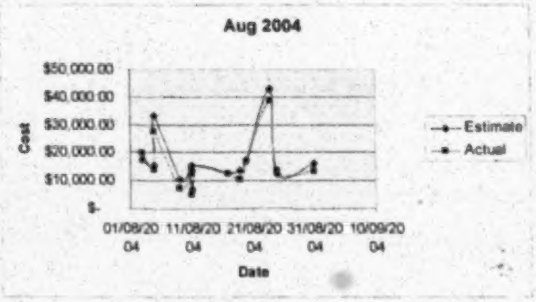

July 1st half 2001
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
Cost
02/07/ 2001
04/07/ 2001
06/07/ 2001
08/07/ 2001
10/07/ 2001
12/07/ 2001
14/07/ 2001
Date
Estimate
Actual

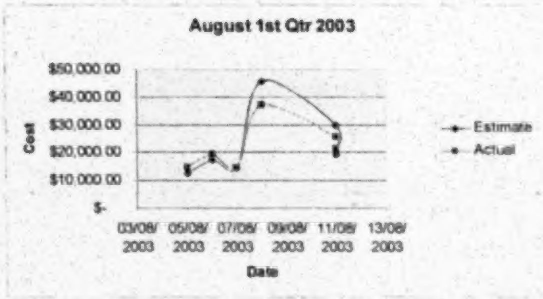

July 2nd half 2001
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
Cost
18/07/2001
23/07/2001
28/07/2001
02/08/2001
Date
Estimate
Actual

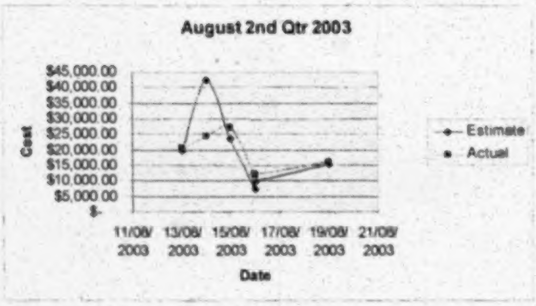

July 1st half 2000
$45,000.00
$40,000.00
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
Cost
03/07/20 00
08/07/20 00
13/07/20 00
18/07/20 00
23/07/20 00
Date
Estimate
Actual

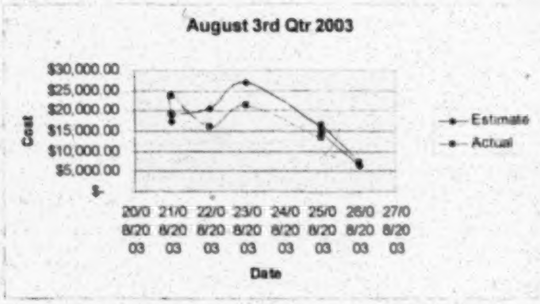

July 2nd half 2000
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
Cost
18/07/2000
23/07/2000
28/07/2000
02/08/2000
Date
Estimate
Actual

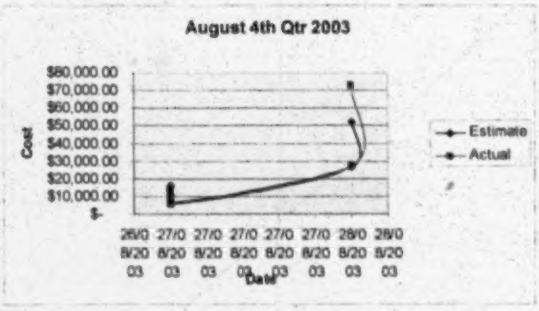

July 1999
$16,000.00
$14,000.00
$12,000.00
$10,000.00
$8,000.00
$6,000.00
$4,000.00
$2,000.00
$-
Cost
22/07/1999
3/07/1999
4/07/1999
25/07/1999
26/07/1999
7/07/1999
28/07/1999
Date
Estimate
Actual

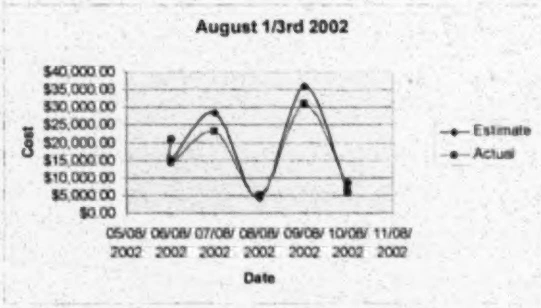

Aug 2004
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$-
Cost
01/08/2004
11/08/2004
21/08/2004
31/08/2004
10/09/2004
Date
Estimate
Actual

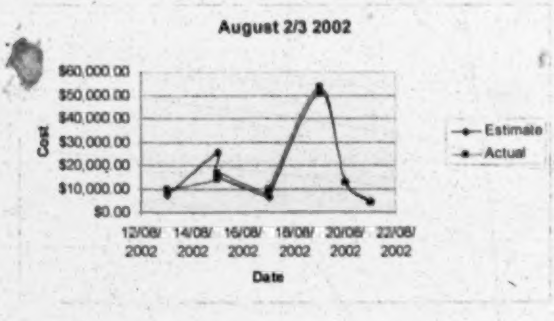

August 1st Qtr 2003
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$-
Cost
03/08/2003
05/08/2003
07/08/2003
09/08/2003
11/08/2003
13/08/2003
Date
Estimate
Actual

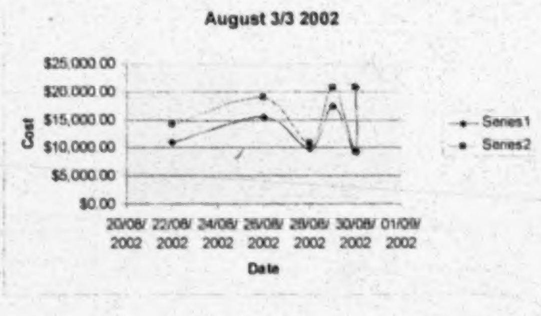

August 2nd Qtr 2003
$45,000.00
$40,000.00
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$-
Cost
11/08/2003
13/08/2003
15/08/2003
17/08/2003
19/08/2003
21/08/2003
Date
Estimate
Actual

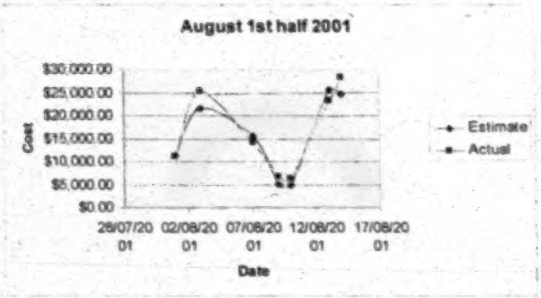

August 3rd Qtr 2003
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$-
Cost
20/08/2003
21/08/2003
22/08/2003
23/08/2003
24/08/2003
25/08/2003
26/08/2003
27/08/2003
Date
Estimate
Actual


August 4th Qtr 2003
$80,000.00
$70,000.00
$60,000.00
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$-
Cost
26/08/2003
27/08/2003
27/08/2003
27/08/2003
27/08/2003
27/08/2003
28/08/2003
28/08/2003
Date
Estimate
Actual

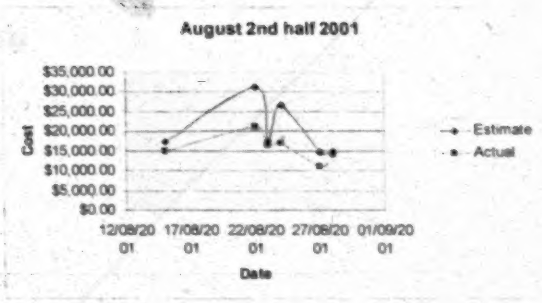

August 1/3rd 2002
$40,000.00
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
Cost
05/08/2002
06/08/2002
07/08/2002
08/08/2002
09/08/2002
10/08/2002
11/08/2002
Date
Estimate
Actual


August 2/3 2002
$60,000.00
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$0.00
Cost
12/08/2002
14/08/2002
16/08/2002
18/08/2002
20/08/2002
22/08/2002
Date
Estimate
Actual

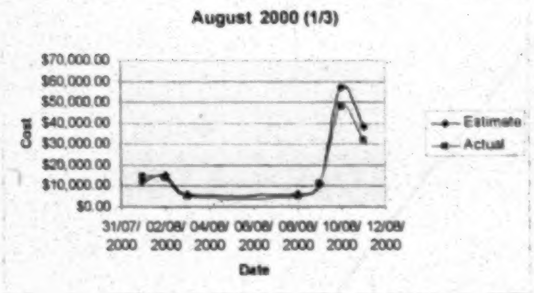

August 3/3 2002
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
Cost
20/08/2002
22/08/2002
24/08/2002
26/08/2002
28/08/2002
30/08/2002
01/09/2002
Date
Series1
Series2



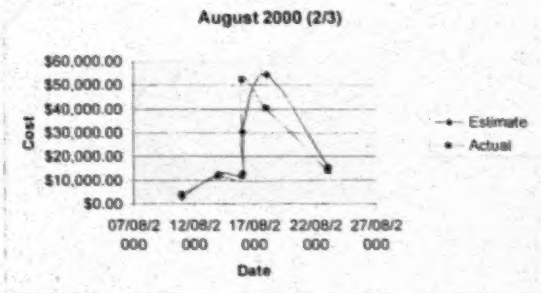
August 1st half 2001



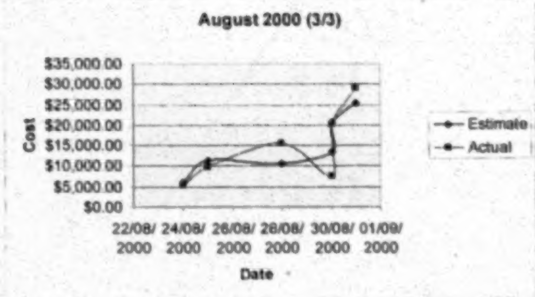
August 2nd half 2001



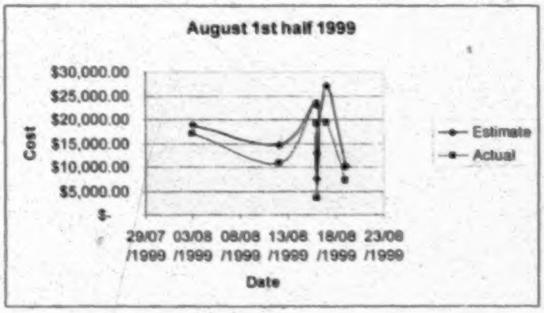
August 2000 (1/3)



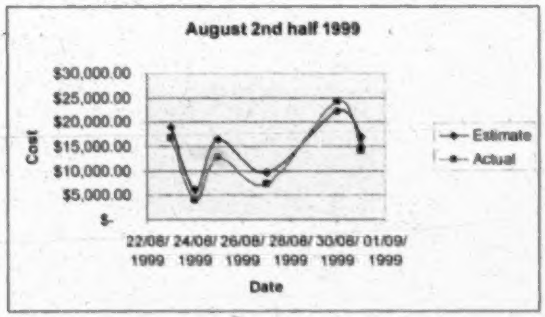
August 2000 (2/3)



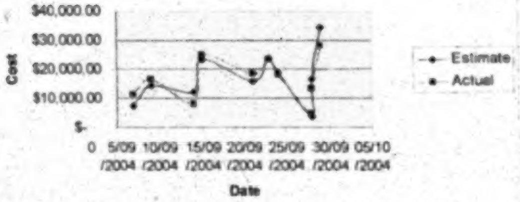
August 2000 (3/3)

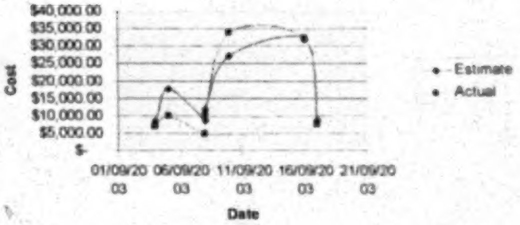


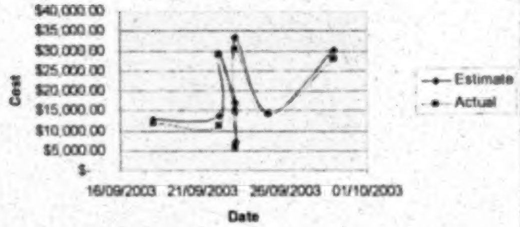
August 1st half 1999

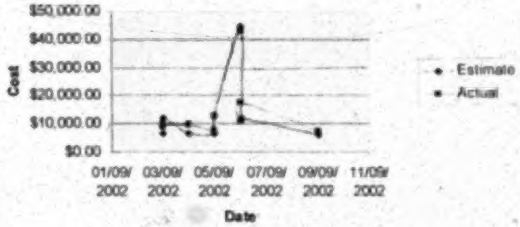


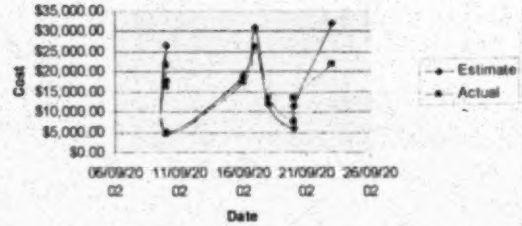
August 2nd half 1999

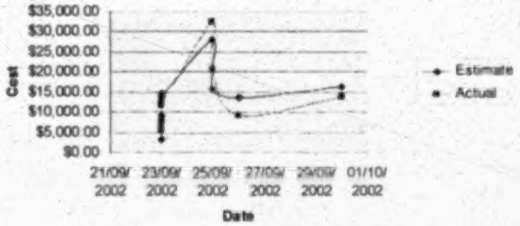

Sept 2004
(Fig 3.5)

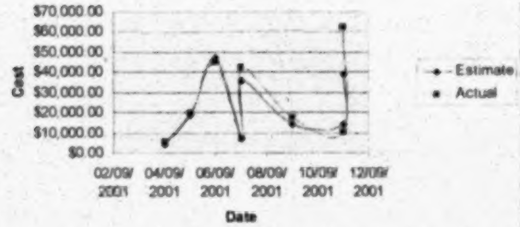

Sept 1st half 2003

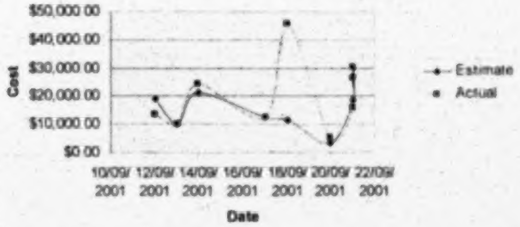

Sept 2nd half 2003

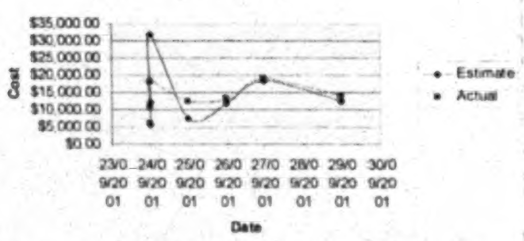

Sept 2002 (1/3)


Sept 2002 (2/3)

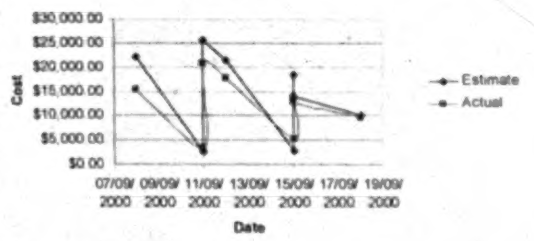

Sept 2002 (3/3)


Sept 2001 (1/3)

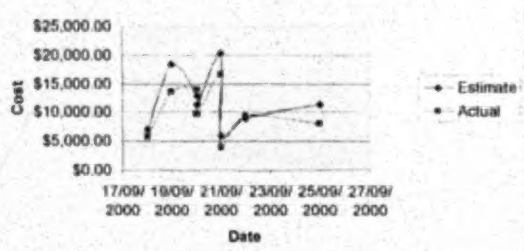

Sept 2001(2/3)



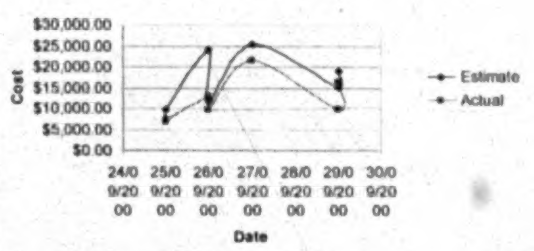
Sept 2001(3/3)



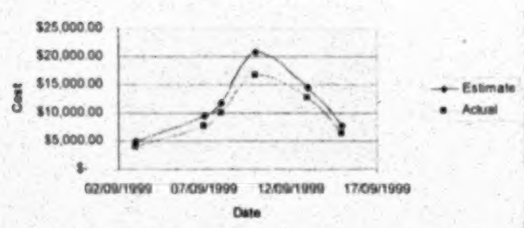
September 2000 (1/3)



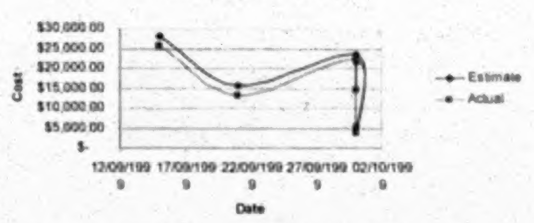
September 2000 (2/3)



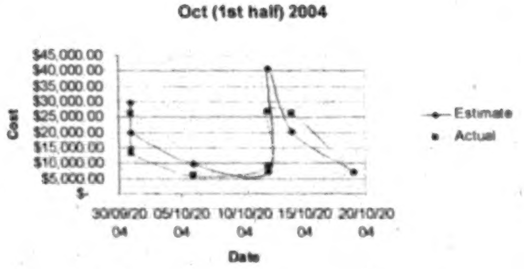
September 2000 (3/3)

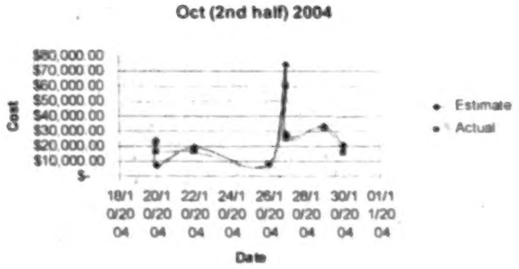


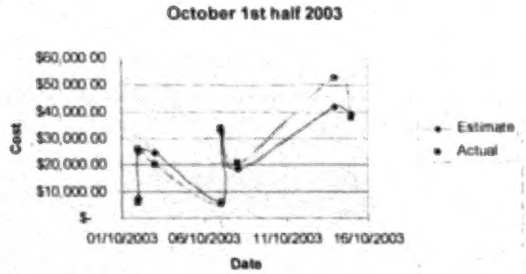
September 1st half 1999

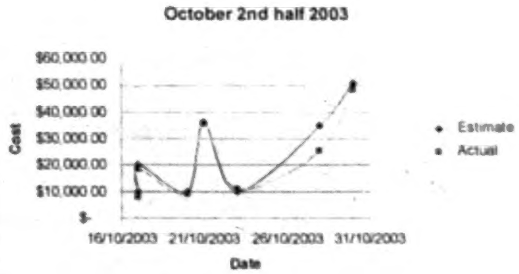


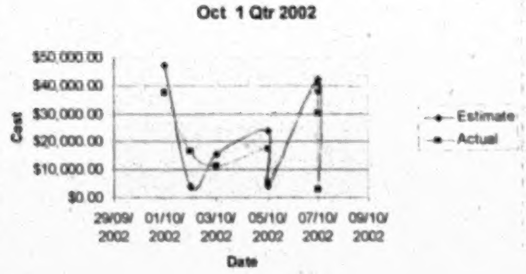
September 2nd half 1999

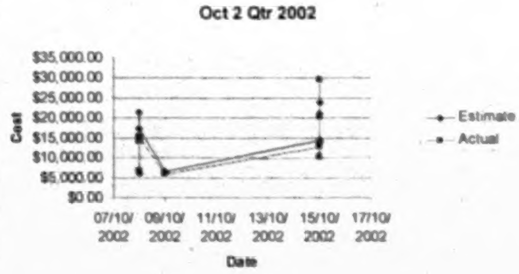

Oct (1st half) 2004

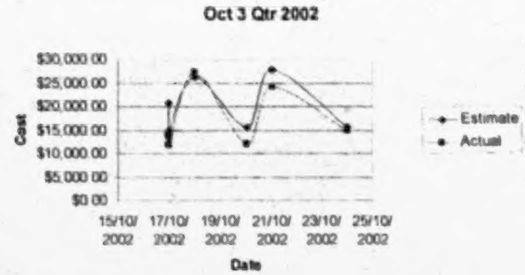

Oct (2nd half) 2004

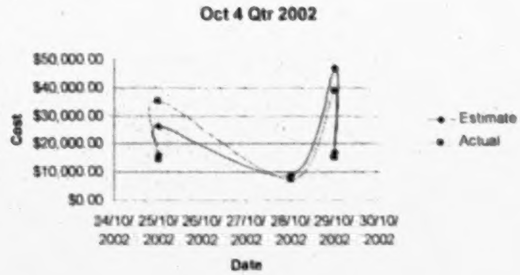

October 1st half 2003

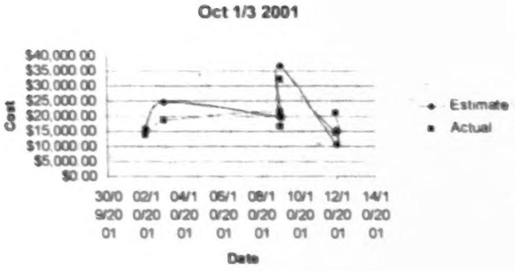

October 2nd half 2003

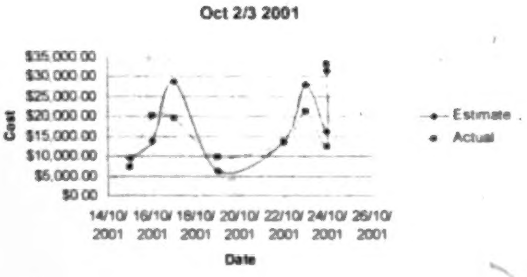

Oct 1 Qtr 2002

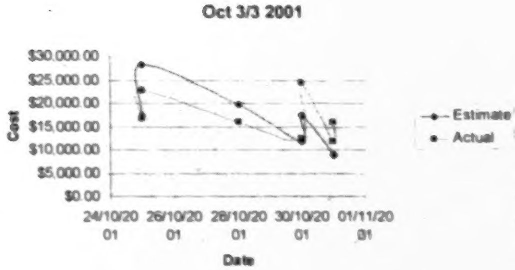

Oct 2 Qtr 2002

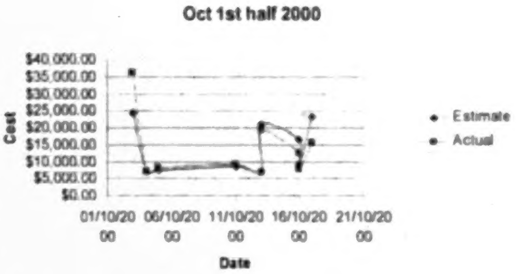

Oct 3 Qtr 2002

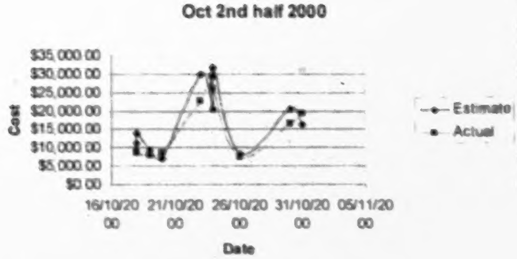

Oct 4 Qtr 2002

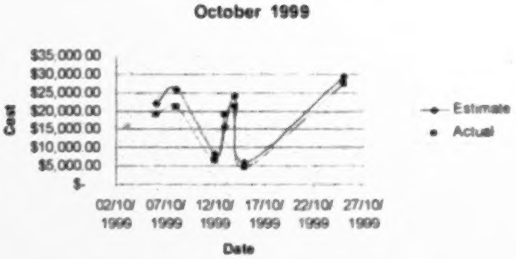

Oct 1/3 2001

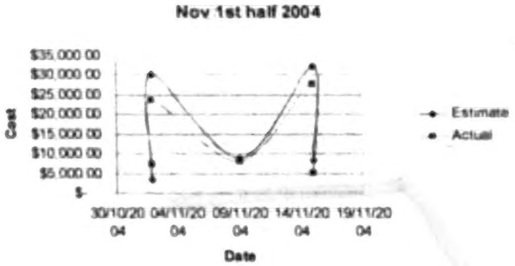

Oct 2/3 2001


Oct 3/3 2001

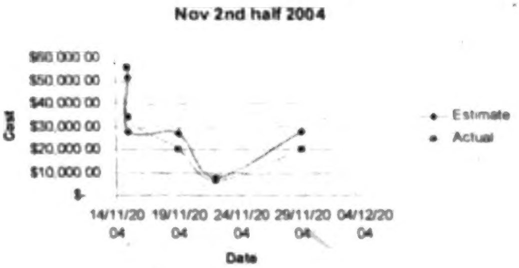

Oct 1st half 2000


Oct 2nd half 2000

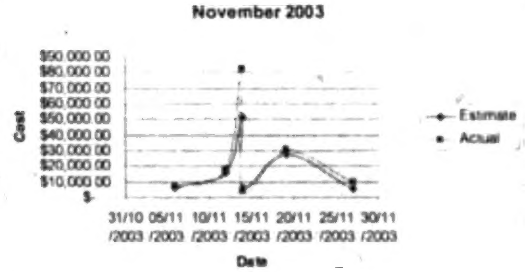

October 1999



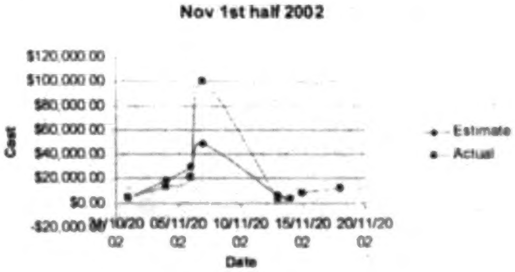
Nov 1st half 2004



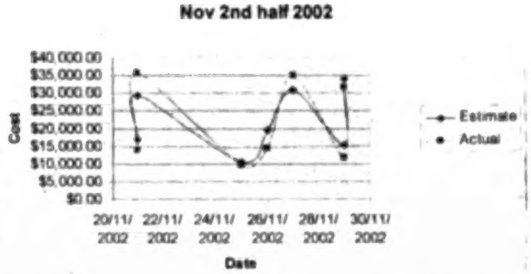
Nov 2nd half 2004



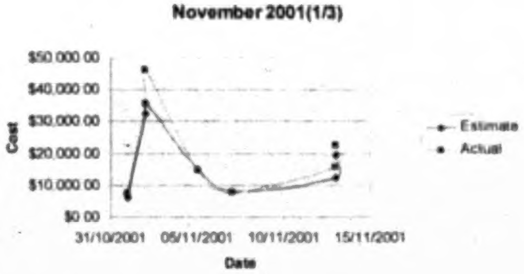
November 2003



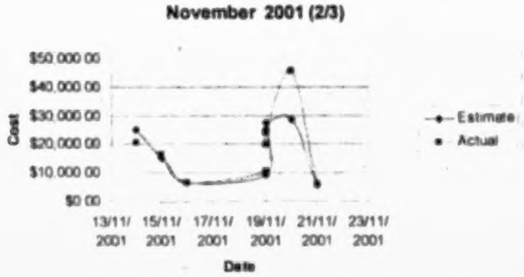
Nov 1st half 2002



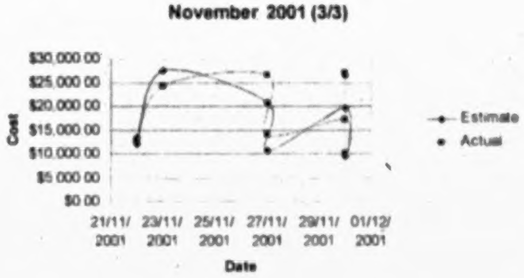
Nov 2nd half 2002



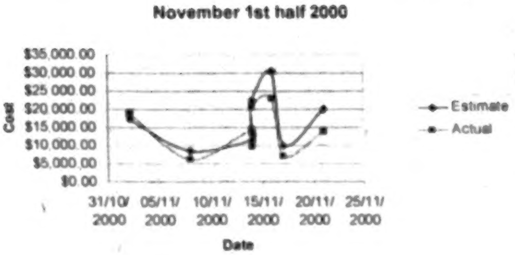
November 2001(1/3)

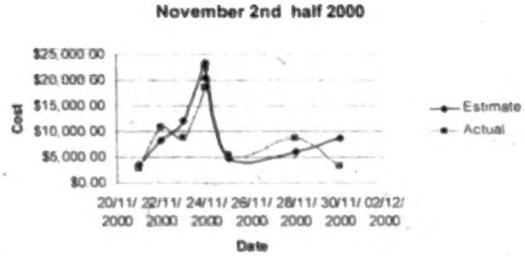


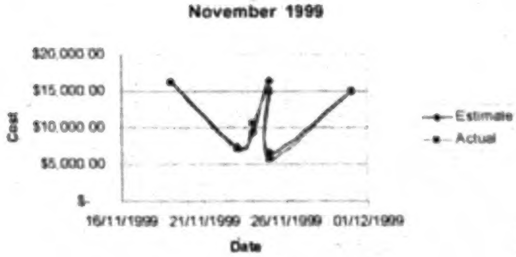
November 2001 (2/3)

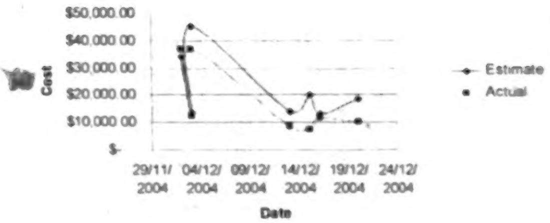


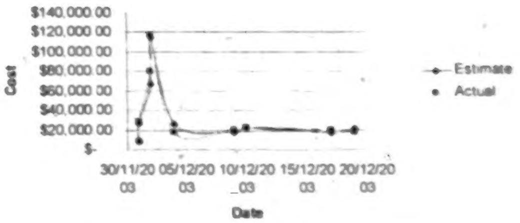
November 2001 (3/3)

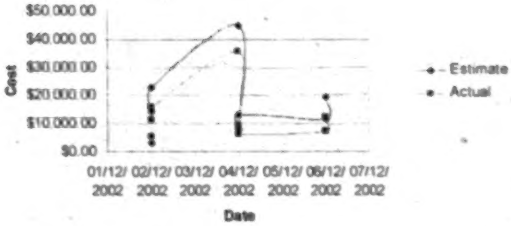

November 1st half 2000
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
Cost
31/10/ 2000
05/11/ 2000
10/11/ 2000
15/11/ 2000
20/11/ 2000
25/11/ 2000
Date
Estimate
Actual

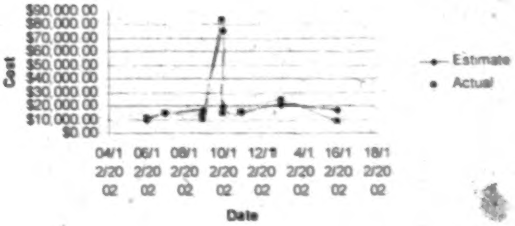

November 2nd half 2000
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
Cost
20/11/ 2000
22/11/ 2000
24/11/ 2000
26/11/ 2000
28/11/ 2000
30/11/ 2000
02/12/ 2000
Date
Estimate
Actual

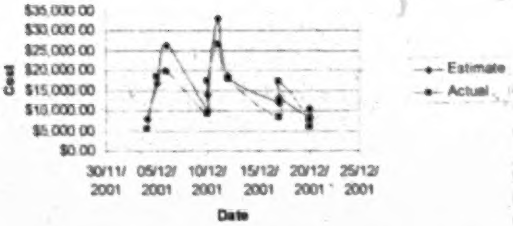

November 1999
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$-
Cost
16/11/1999
21/11/1999
26/11/1999
01/12/1999
Date
Estimate
Actual

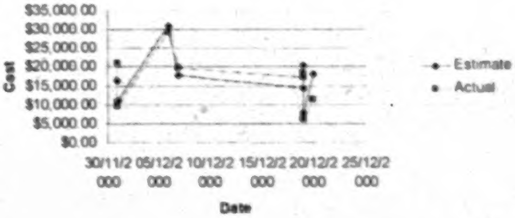

Dec 2004

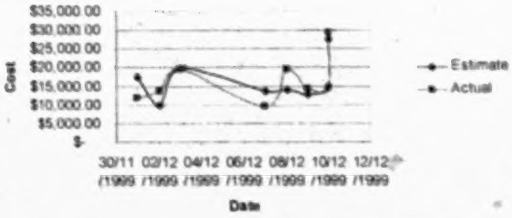

December 2003

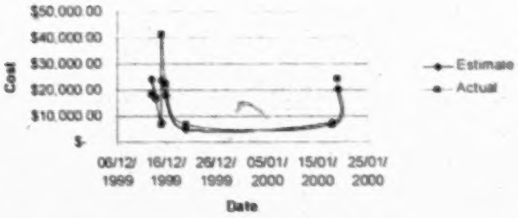

Dec 1st half 2002


Dec 2nd half 2002

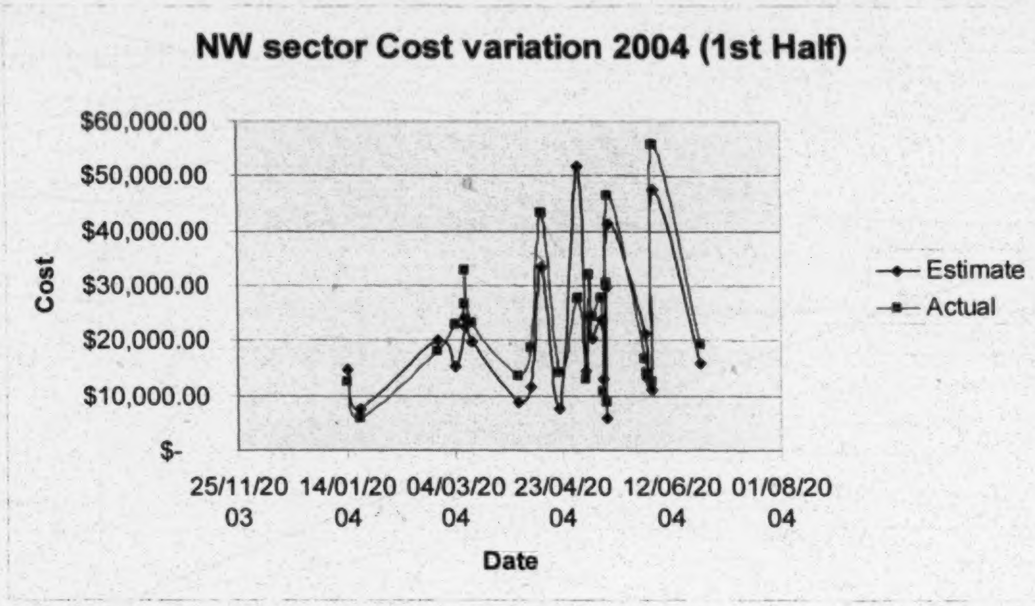

Dec 2001

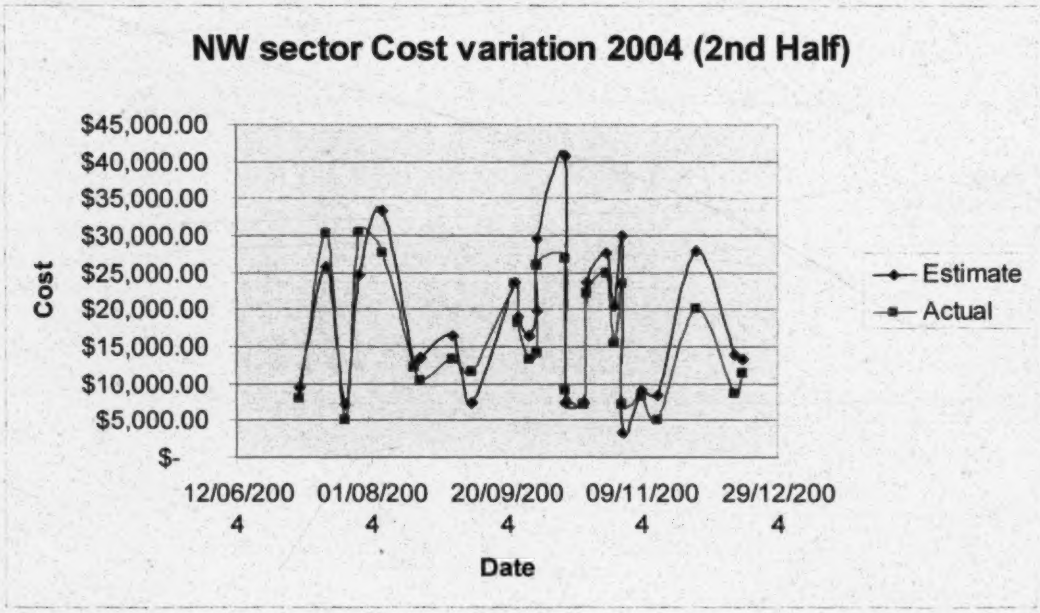

December 2000

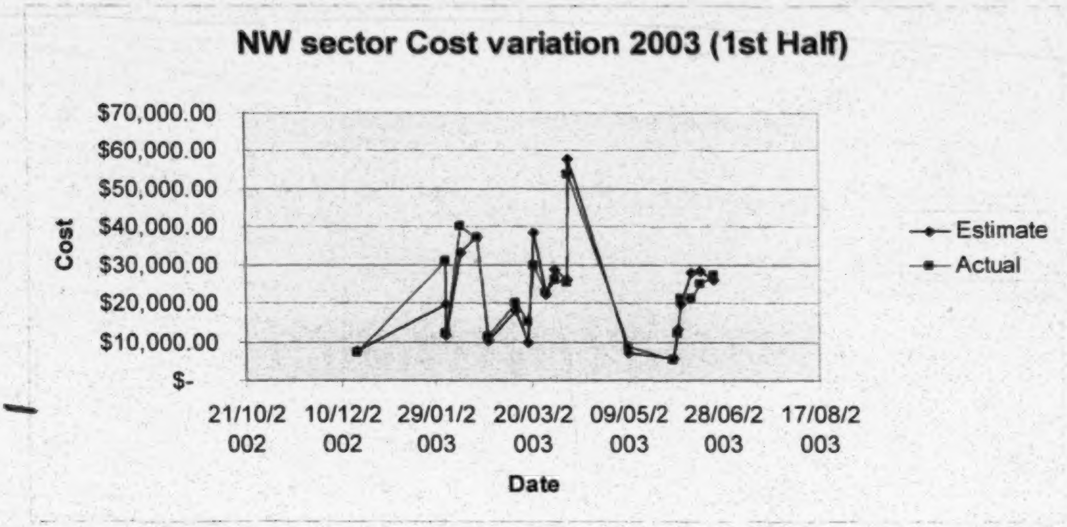

December 1st half 1999

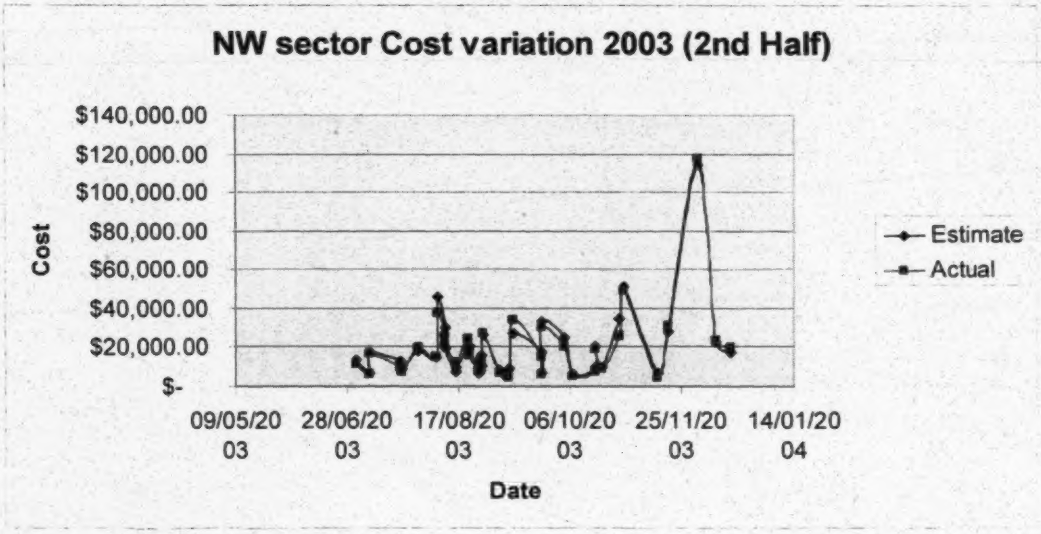

December 2nd half 1999

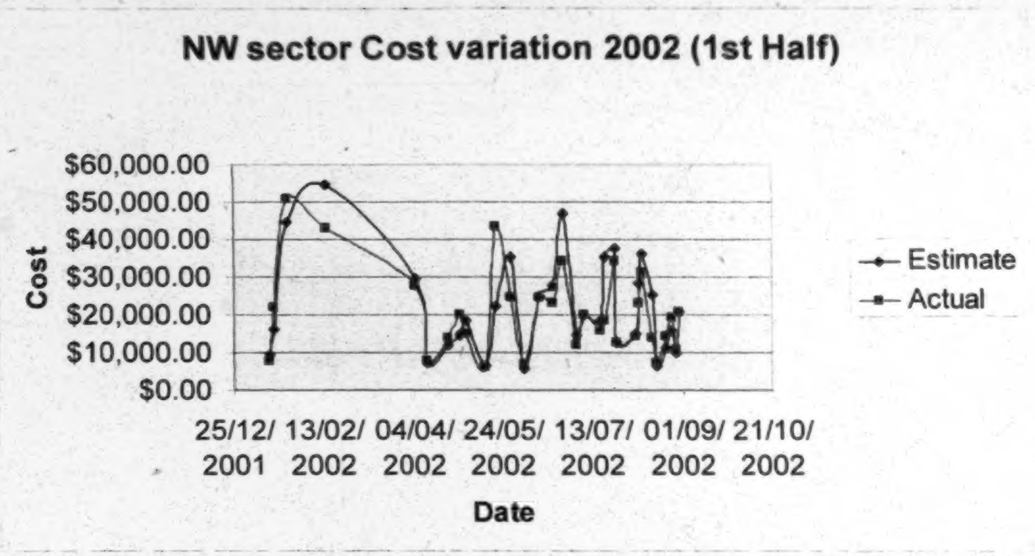
Appendix C

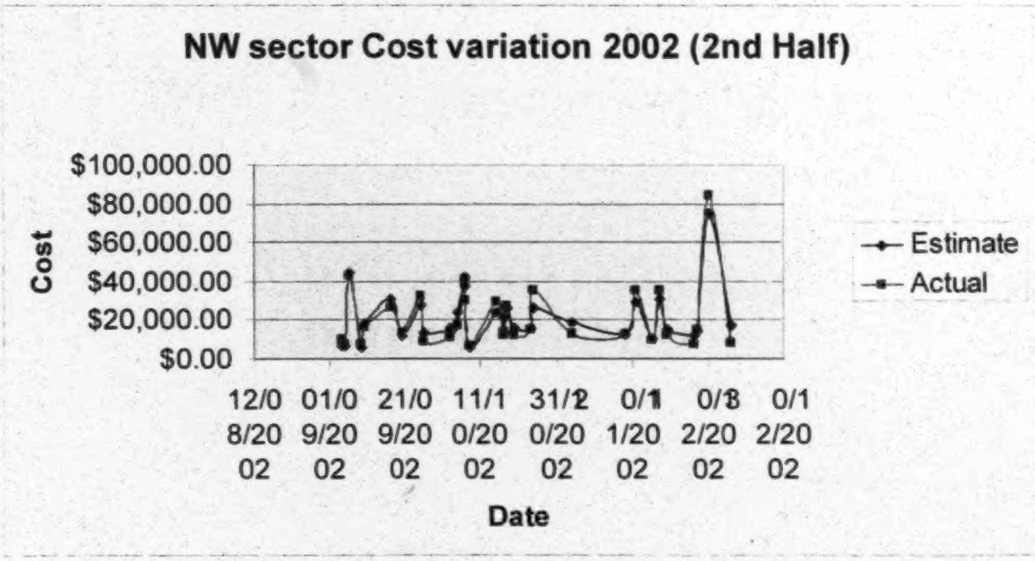

NW sector Cost variation 2004 (1st Half)
$60,000.00
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$-
Cost
25/11/2003
14/01/2004
04/03/2004
23/04/2004
12/06/2004
01/08/2004
Date
Estimate
Actual

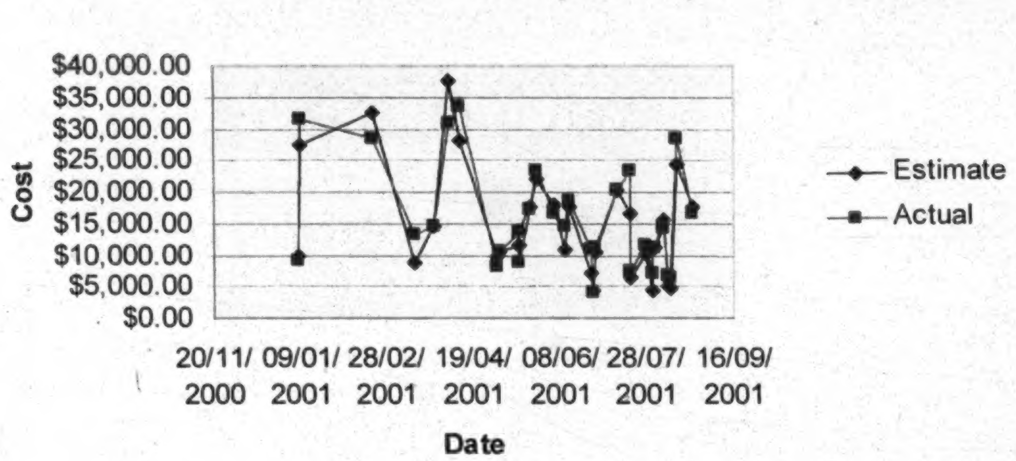

NW sector Cost variation 2004 (2nd Half)
$45,000.00
$40,000.00
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$-
Cost
12/06/2004
01/08/2004
20/09/2004
09/11/2004
29/12/2004
Date
Estimate
Actual

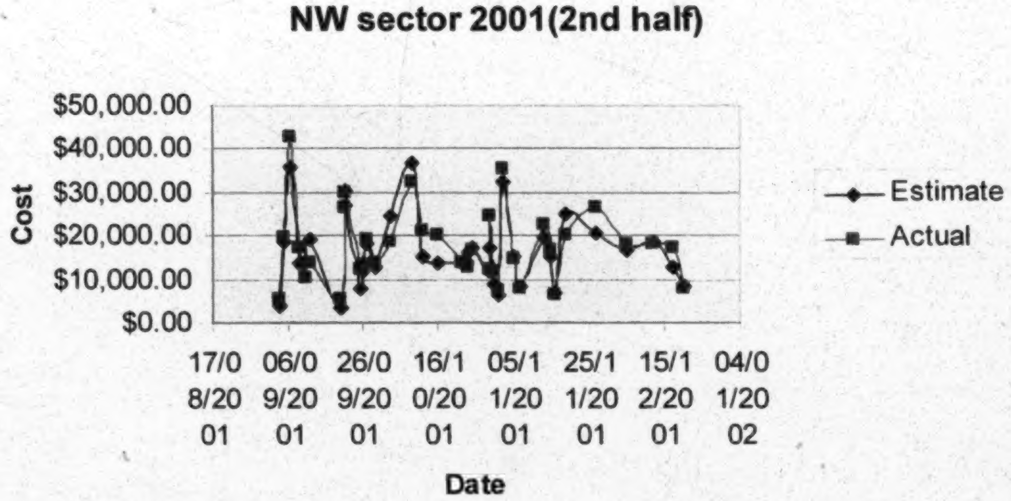

NW sector Cost variation 2003 (1st Half)

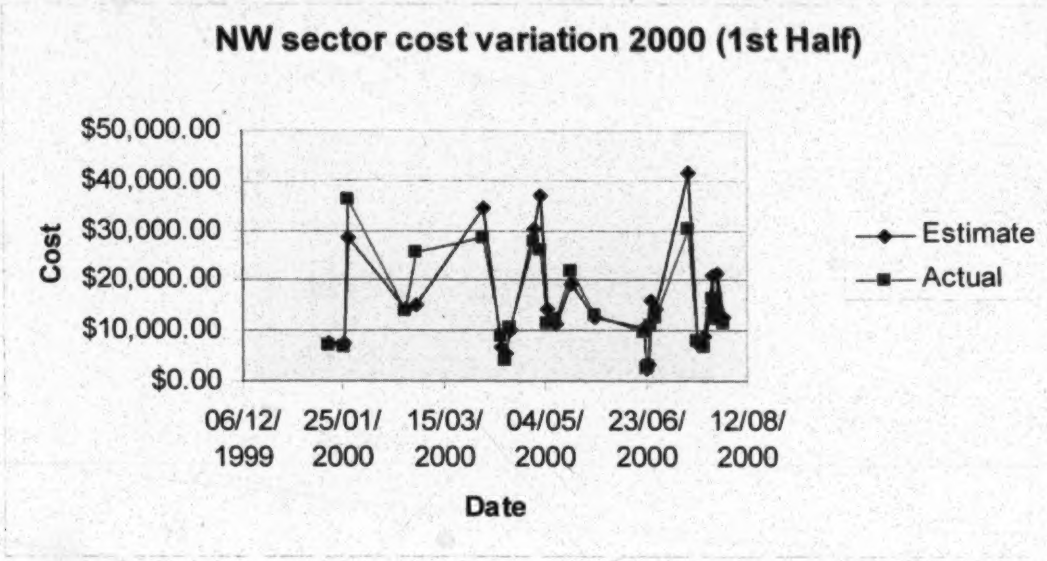

NW sector Cost variation 2003 (2nd Half)

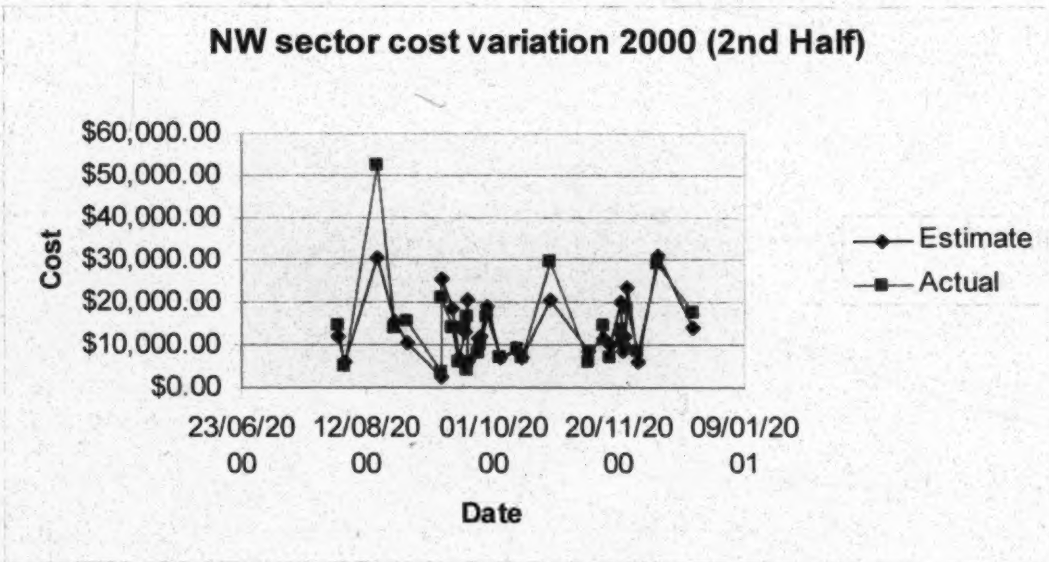

NW sector Cost variation 2002 (1st Half)
Cost
$60,000.00
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$0.00
25/12/2001
13/02/2002
04/04/2002
24/05/2002
13/07/2002
01/09/2002
21/10/2002
Date
Estimate
Actual

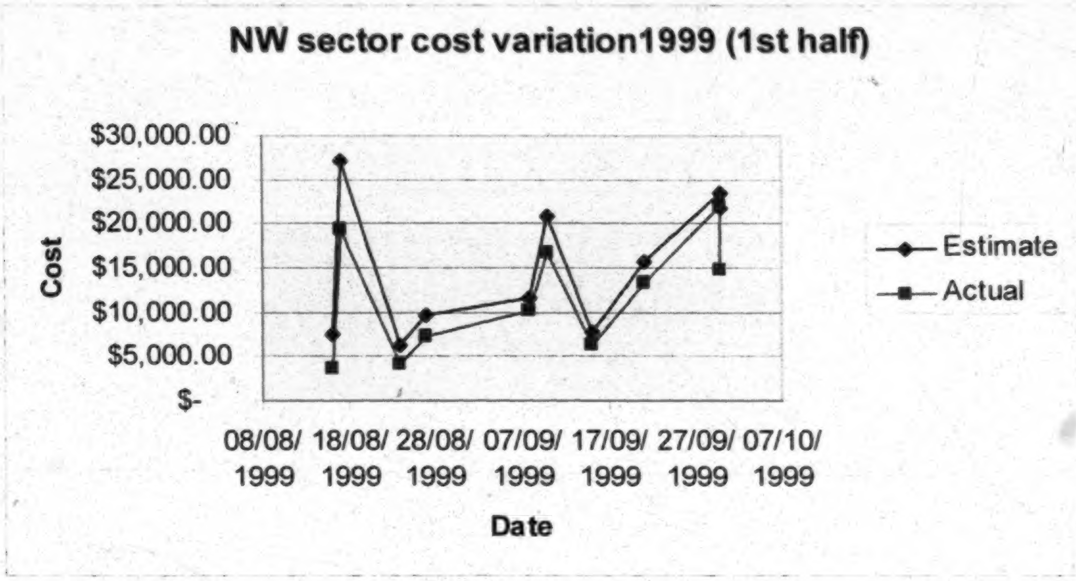

NW sector Cost variation 2002 (2nd Half)
Cost
$100,000.00
$80,000.00
$60,000.00
$40,000.00
$20,000.00
$0.00
12/08/2002
01/09/2002
21/09/2002
11/10/2002
31/10/2002
0/11/2002
0/12/2002
0/12/2002
Date
Estimate
Actual

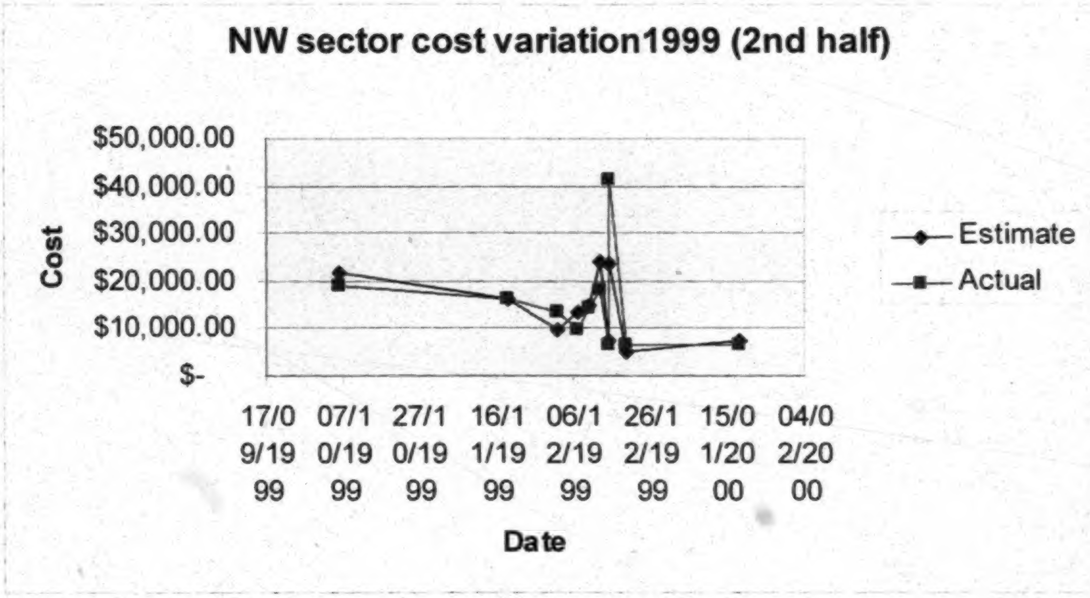

NW sector 2001(1st half)
$40,000.00
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
Cost
20/11/2000
09/01/2001
28/02/2001
19/04/2001
08/06/2001
28/07/2001
16/09/2001
Date
Estimate
Actual

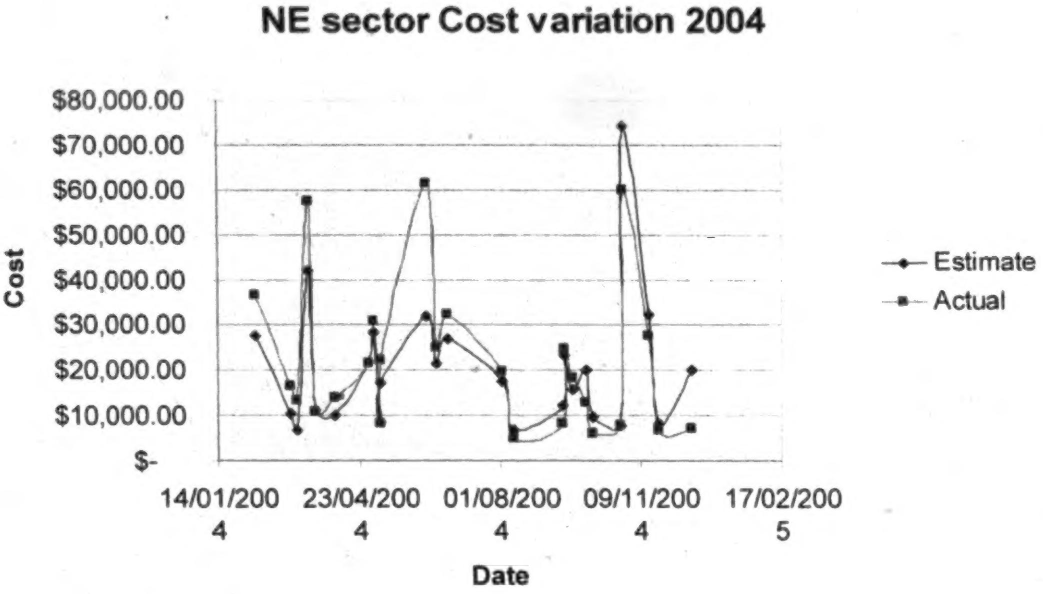

NW sector 2001(2nd half)
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$0.00
Cost
17/08/2001
06/09/2001
26/09/2001
16/10/2001
05/11/2001
25/11/2001
15/12/2001
04/01/2002
Date
Estimate
Actual

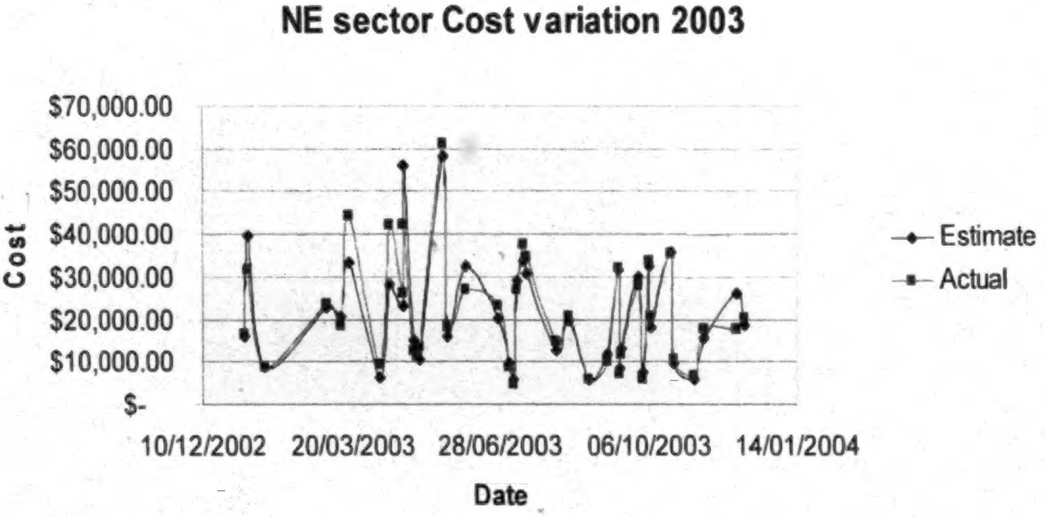

NW sector cost variation 2000 (1st Half)
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$0.00
Cost
06/12/1999
25/01/2000
15/03/2000
04/05/2000
23/06/2000
12/08/2000
Date
Estimate
Actual

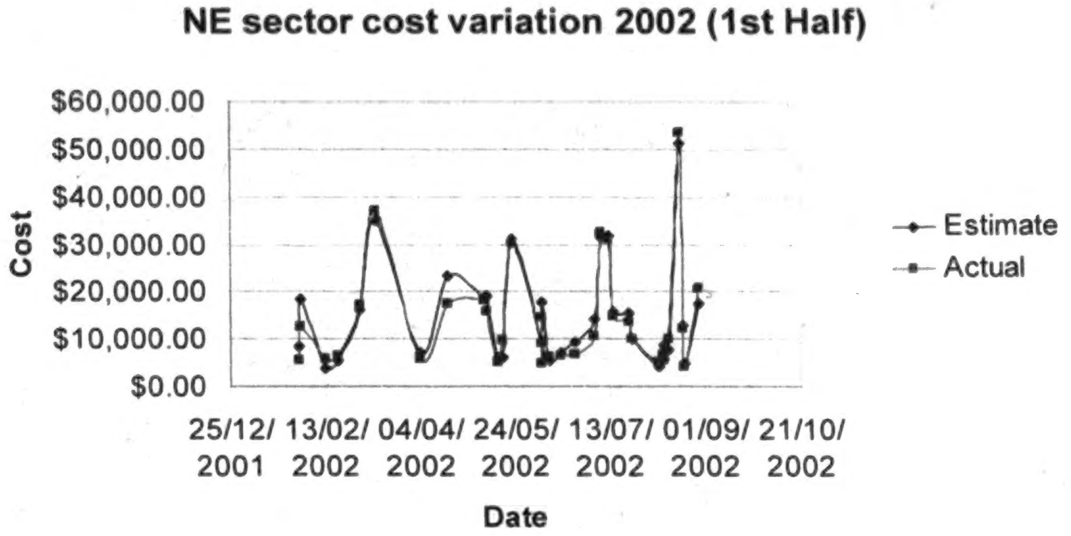

NW sector cost variation 2000 (2nd Half)
$60,000.00
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$0.00
Cost
23/06/2000
12/08/2000
01/10/2000
20/11/2000
09/01/2001
Date
Estimate
Actual

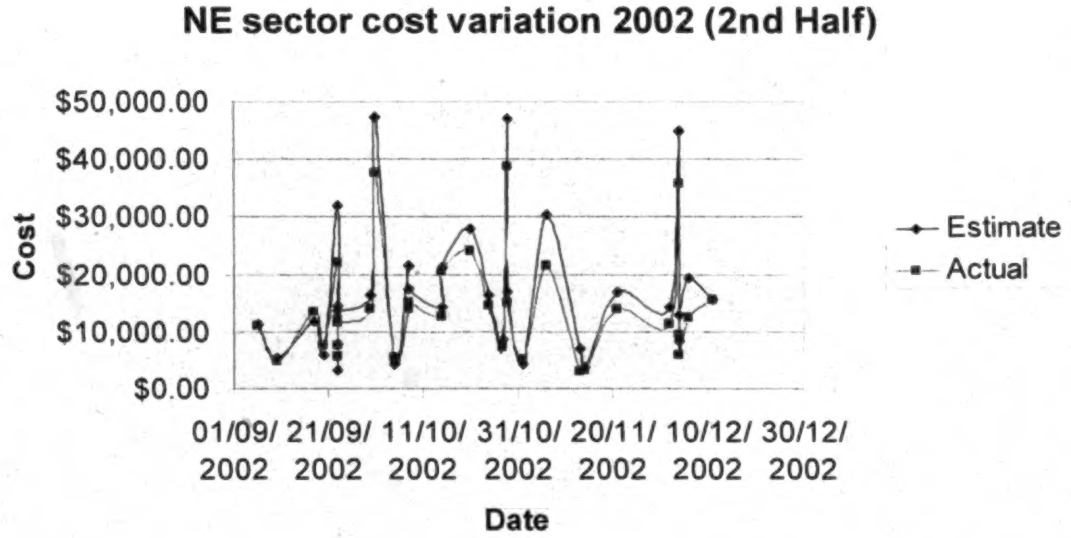

NW sector cost variation1999 (1st half)
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$-
Cost
08/08/ 1999
18/08/ 1999
28/08/ 1999
07/09/ 1999
17/09/ 1999
27/09/ 1999
07/10/ 1999
Date
Estimate
Actual


NW sector cost variation1999 (2nd half)
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$-
Cost
17/09/1999
07/10/1999
27/10/1999
16/11/1999
06/12/1999
26/12/1999
15/01/2000
04/02/2000
Date
Estimate
Actual

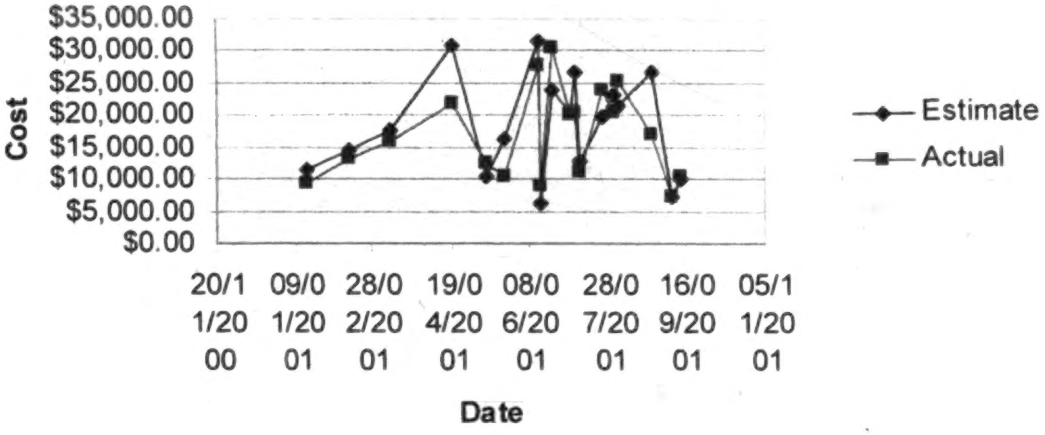

NE sector Cost variation 2004
Cost
$80,000.00
$70,000.00
$60,000.00
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$-
14/01/2004
23/04/2004
01/08/2004
09/11/2004
17/02/2005
Date
Estimate
Actual


NE sector Cost variation 2003
Cost
$70,000.00
$60,000.00
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$-
10/12/2002
20/03/2003
28/06/2003
06/10/2003
14/01/2004
Date
Estimate
Actual

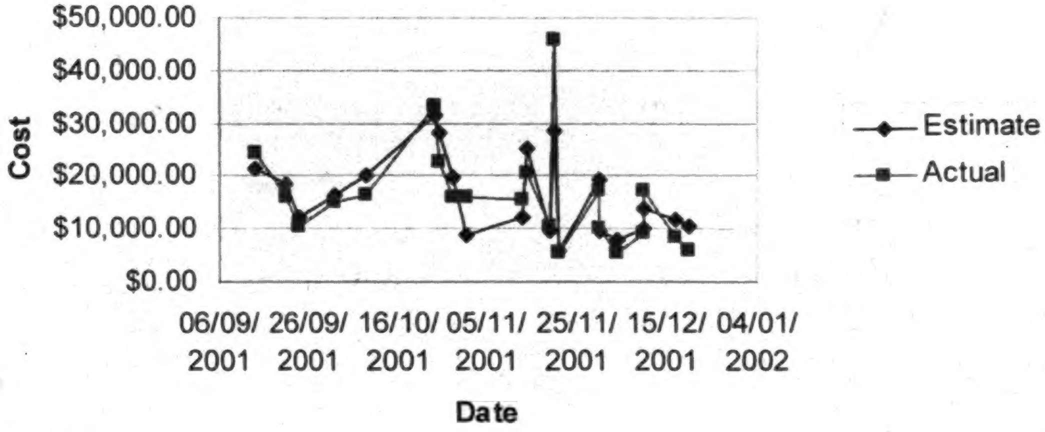

NE sector cost variation 2002 (1st Half)
Cost
$60,000.00
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$0.00
25/12/2001
13/02/2002
04/04/2002
24/05/2002
13/07/2002
01/09/2002
21/10/2002
Date
Estimate
Actual


NE sector cost variation 2002 (2nd Half)
Cost
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$0.00
01/09/2002
21/09/2002
11/10/2002
31/10/2002
20/11/2002
10/12/2002
30/12/2002
Date
Estimate
Actual

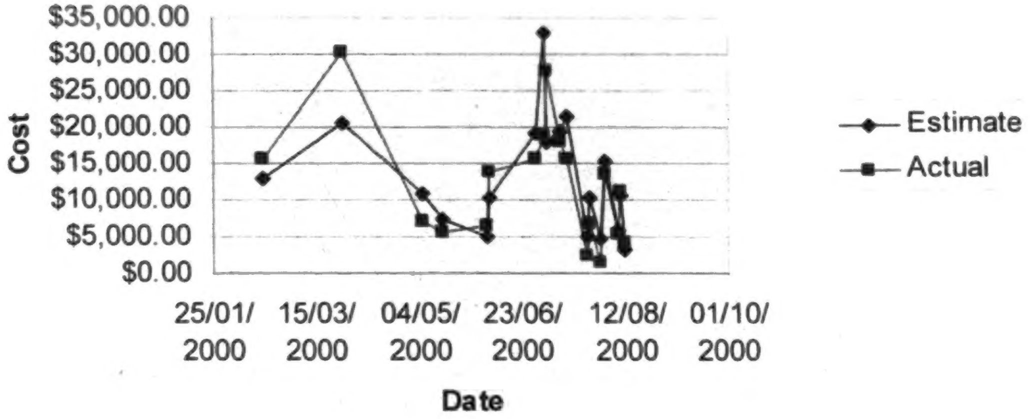
NE sector cost variation 2001 (1st Half)



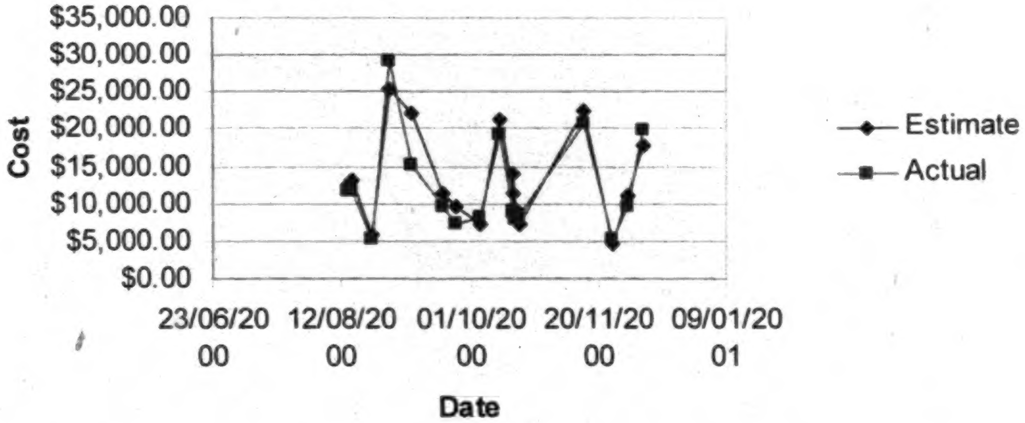
NE sector cost variation 2001 (2nd Half)

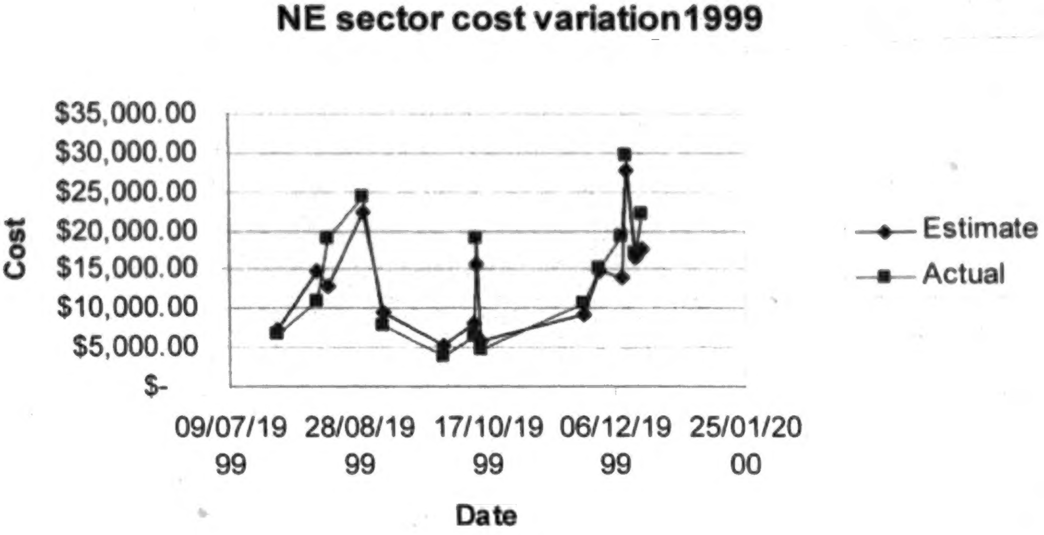


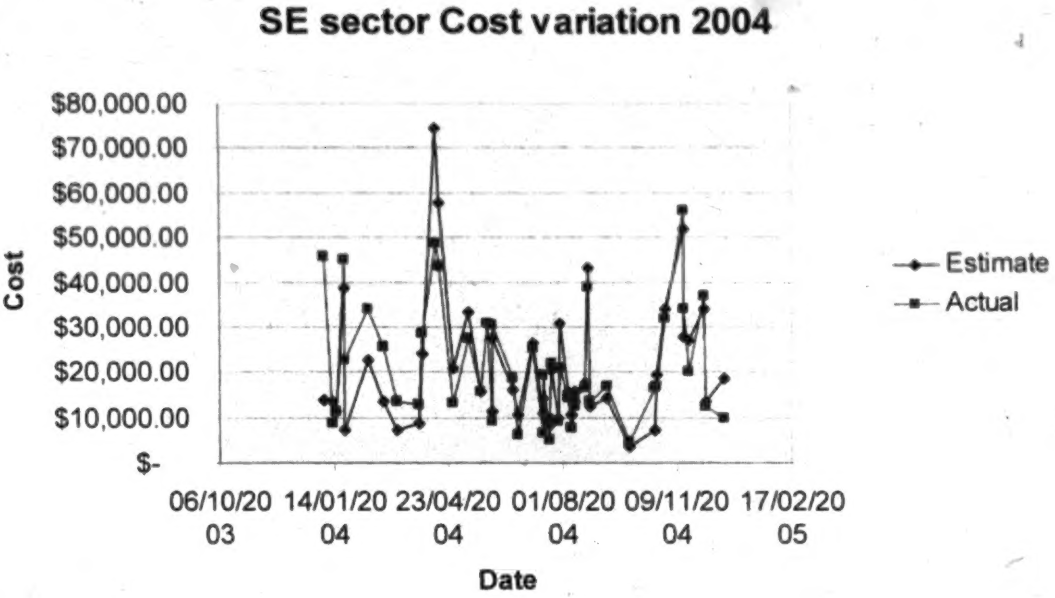
NE sector cost variation 2000 (1st Half)

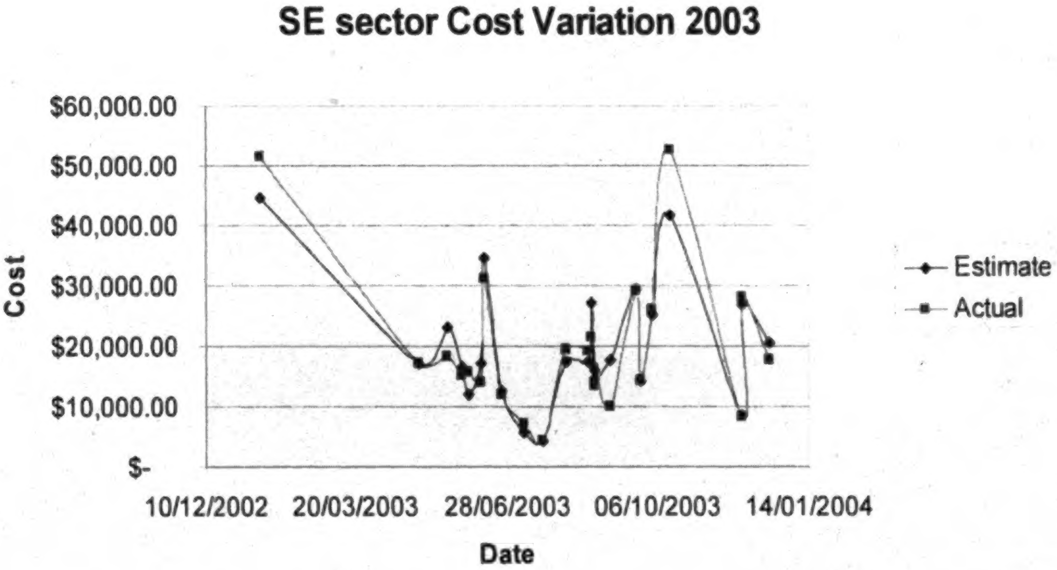


NE sector cost variation 2000 (2nd Half)

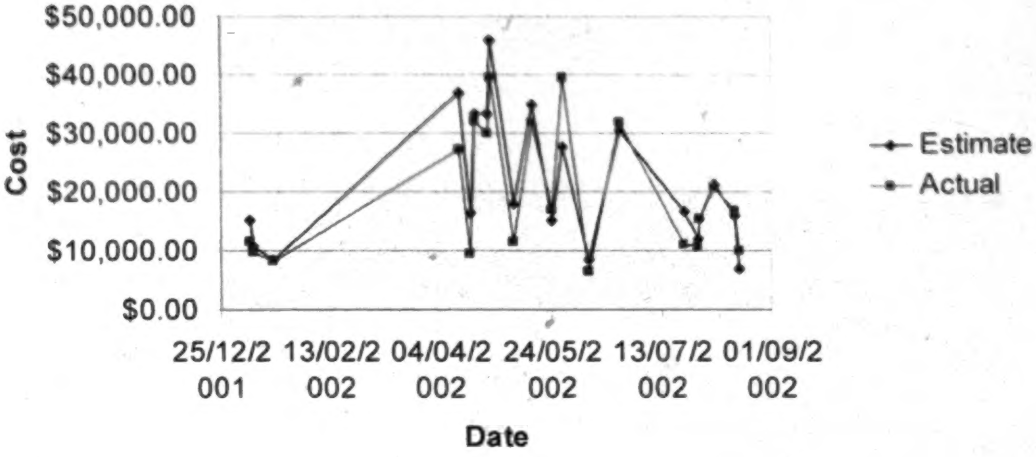



NE sector cost variation1999
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$-
Cost
09/07/1999
28/08/1999
17/10/1999
06/12/1999
25/01/2000
Date
Estimate
Actual

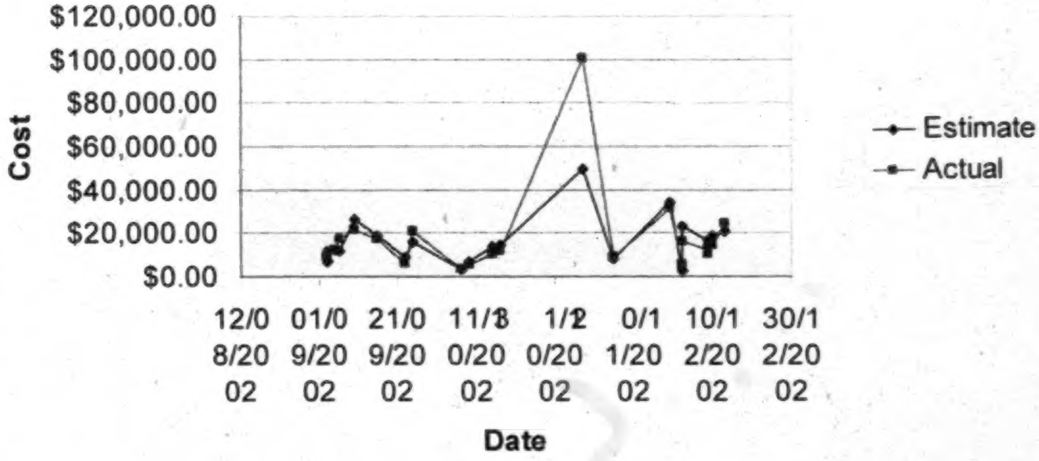

SE sector Cost variation 2004
$80,000.00
$70,000.00
$60,000.00
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$-
Cost
06/10/2003
14/01/2004
23/04/2004
01/08/2004
09/11/2004
17/02/2005
Date
Estimate
Actual

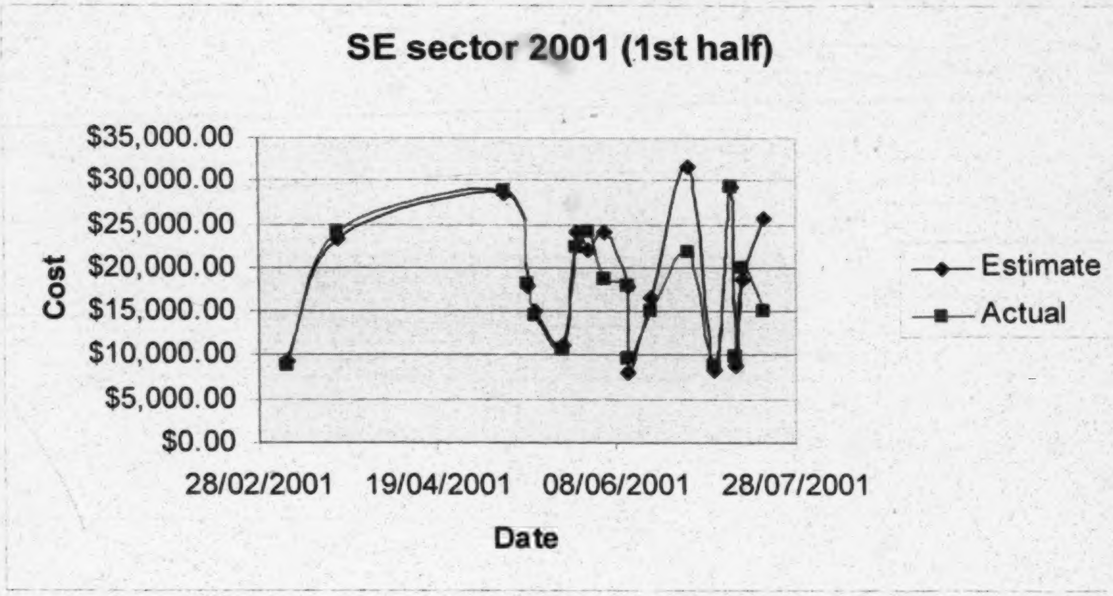

SE sector Cost Variation 2003
$60,000.00
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$-
Cost
10/12/2002
20/03/2003
28/06/2003
06/10/2003
14/01/2004
Date
Estimate
Actual

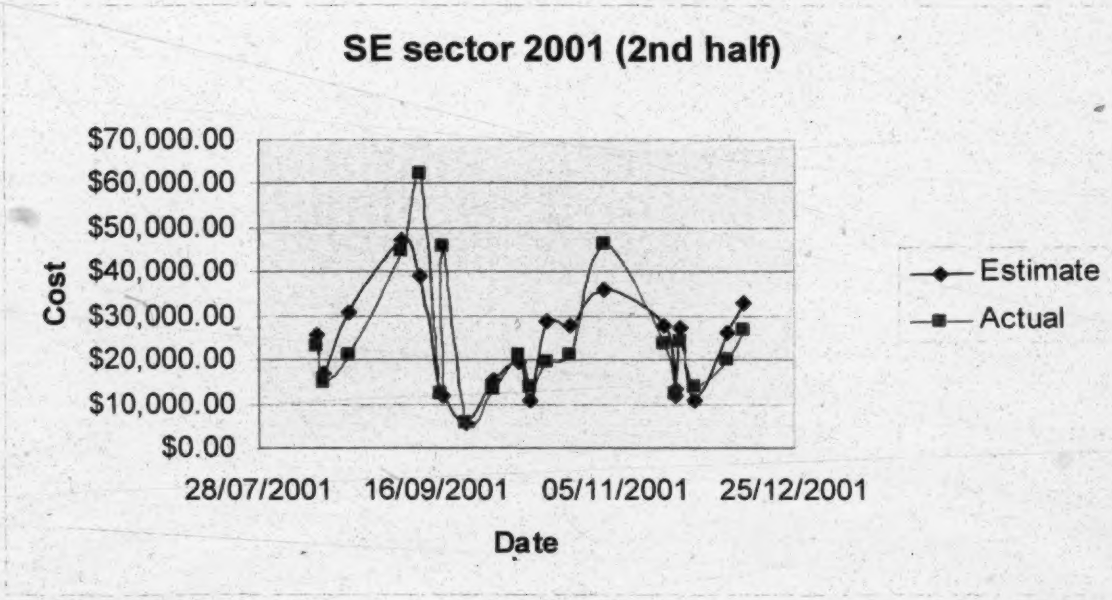
SE sector Cost variation 2002 (1st half)

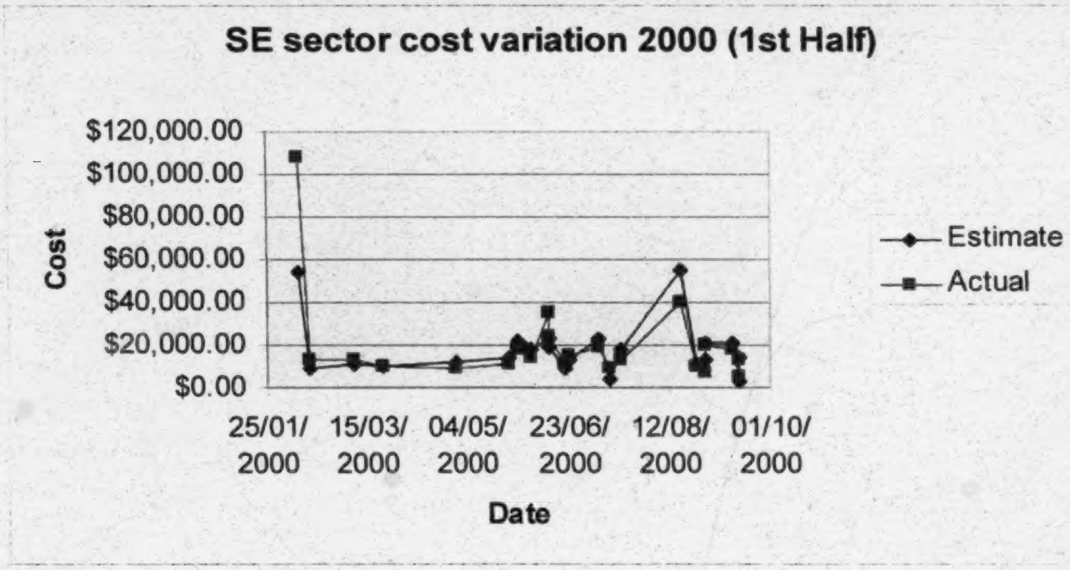


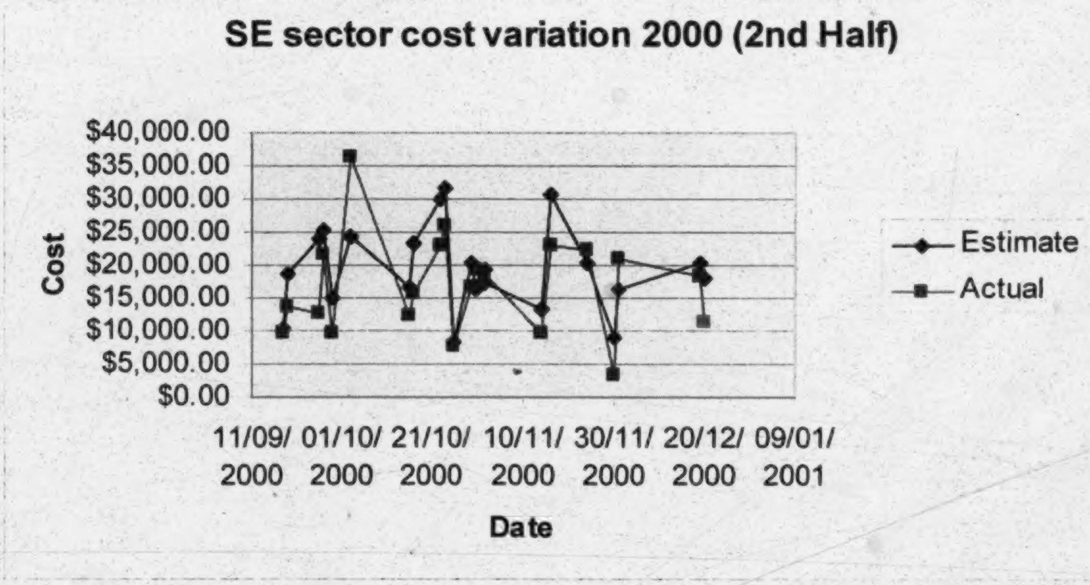
SE sector Cost variation 2002 (2nd half)

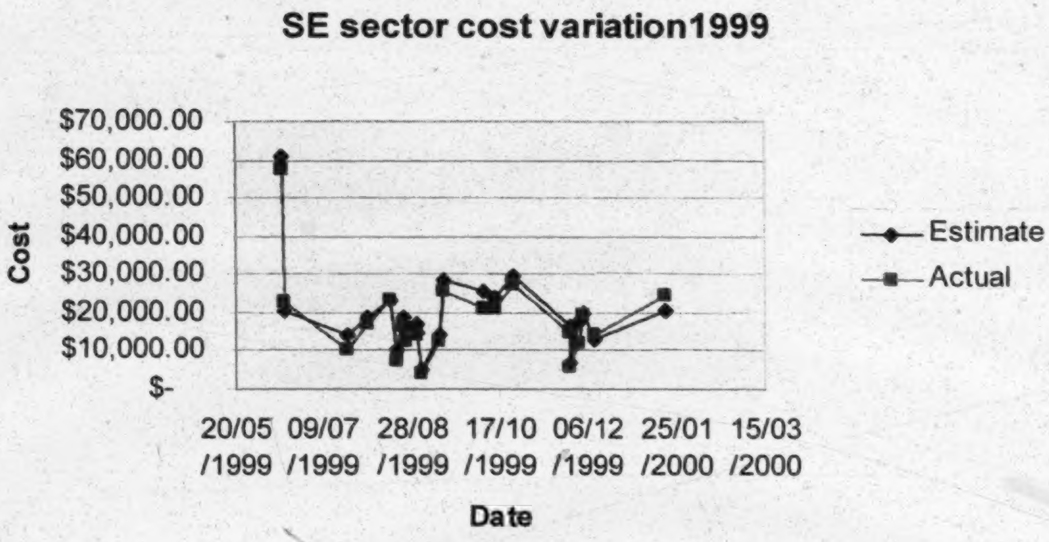



SE sector 2001 (1st half)
Cost
$35,000.00
$30,000.00
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
28/02/2001
19/04/2001
08/06/2001
28/07/2001
Date
Estimate
Actual

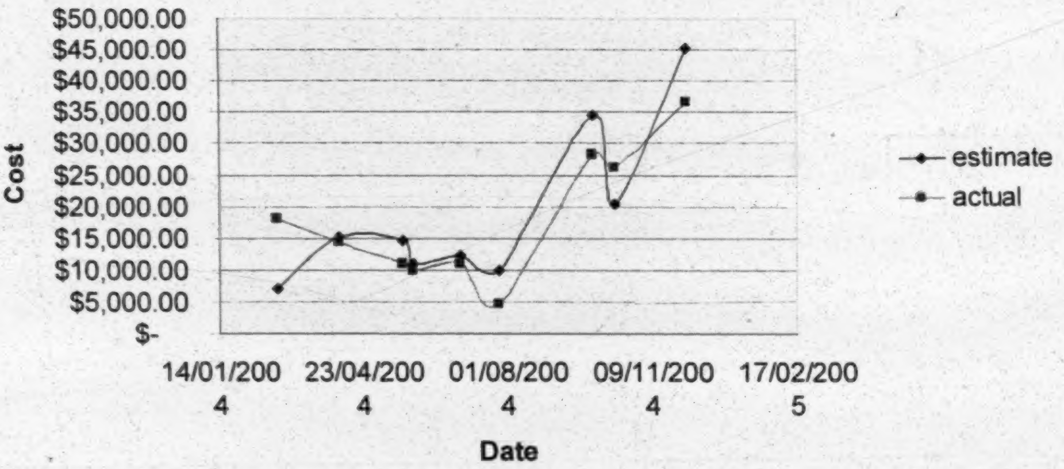

SE sector 2001 (2nd half)
Cost
$70,000.00
$60,000.00
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$0.00
28/07/2001
16/09/2001
05/11/2001
25/12/2001
Date
Estimate
Actual


SE sector cost variation 2000 (1st Half)

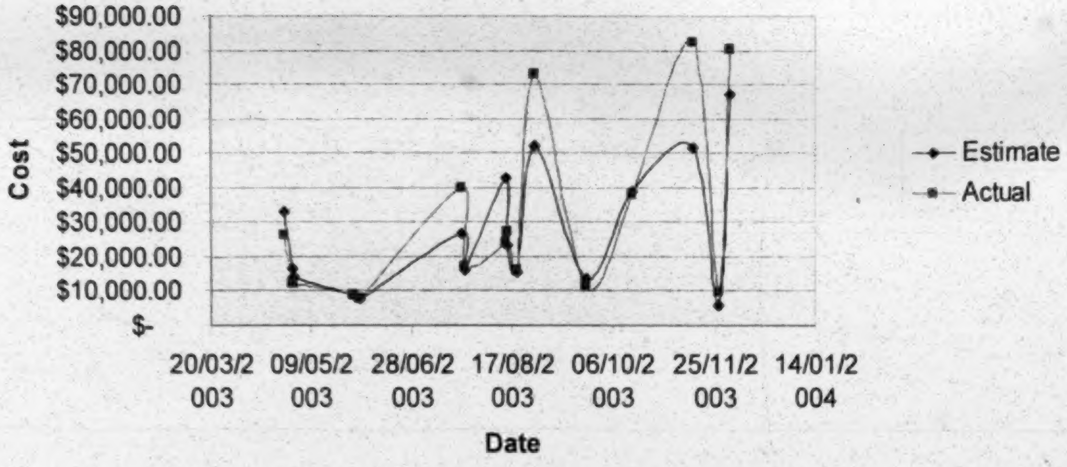

SE sector cost variation 2000 (2nd Half)

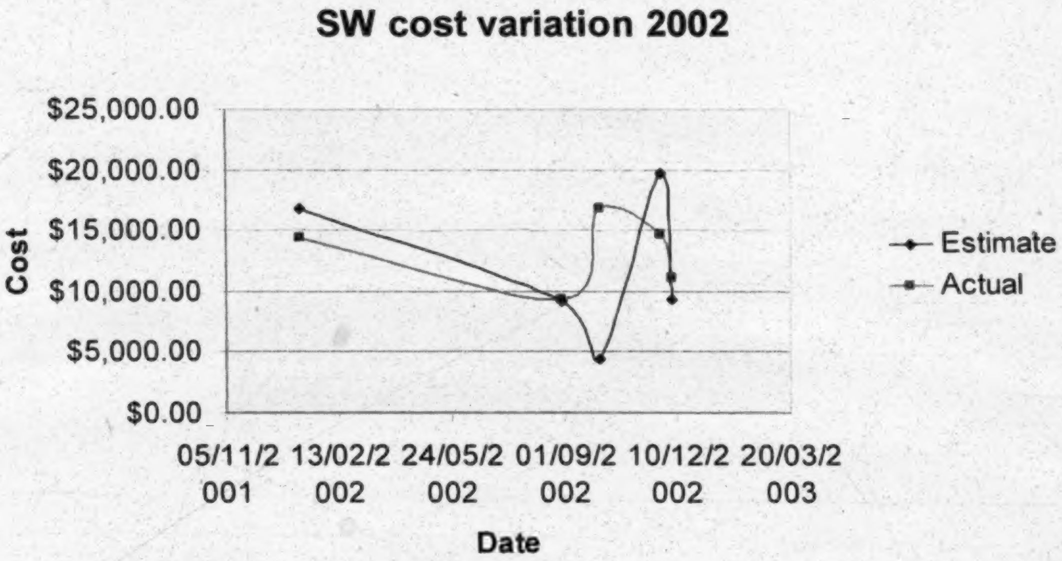

SE sector cost variation1999
Cost
$70,000.00
$60,000.00
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$-
20/05/1999
09/07/1999
28/08/1999
17/10/1999
06/12/1999
25/01/2000
15/03/2000
Date
Estimate
Actual

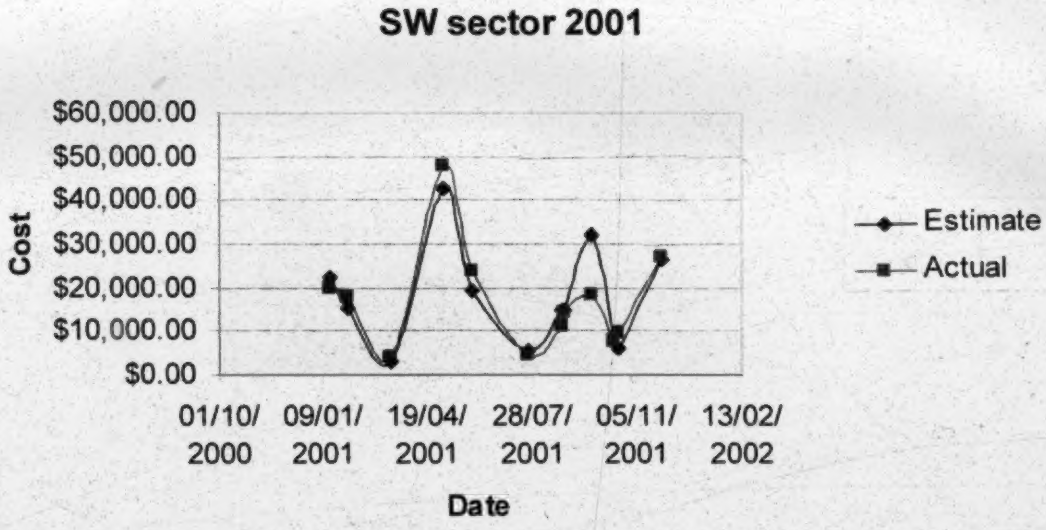
SW sector Cost variation 2004

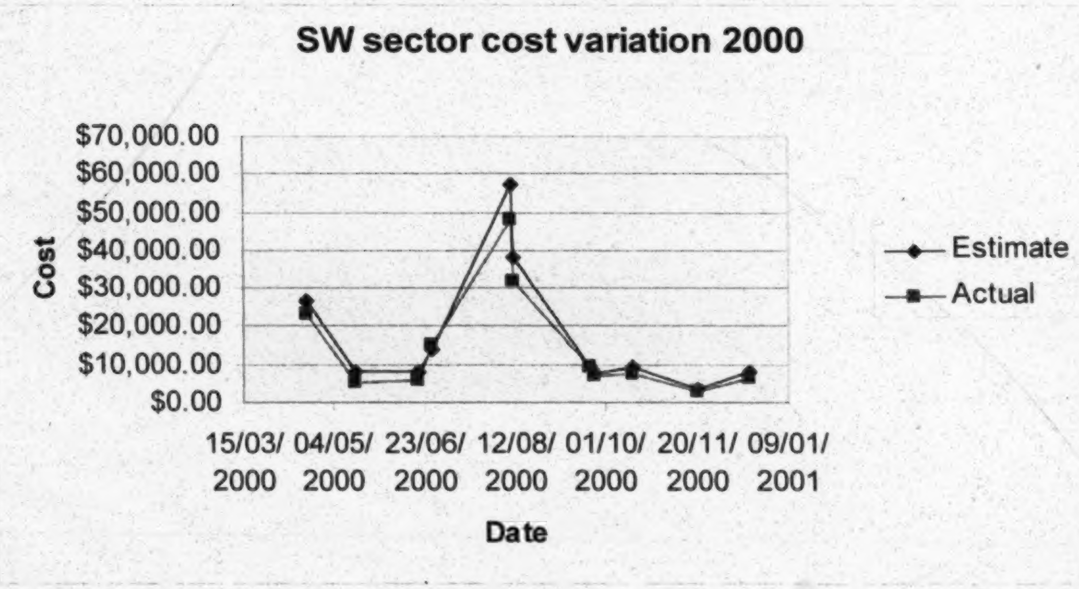


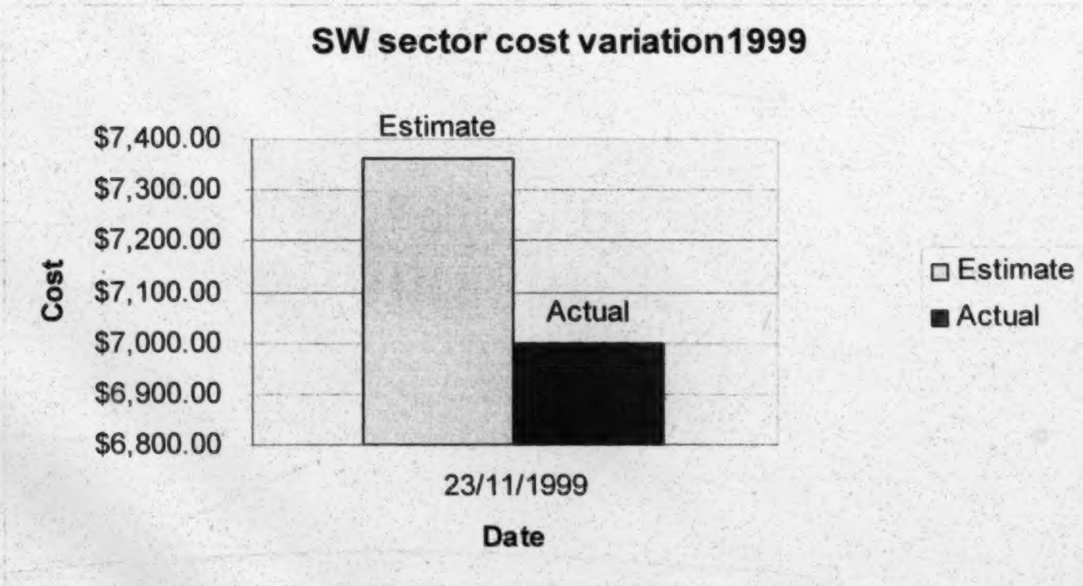
SW sector cost variation 2003

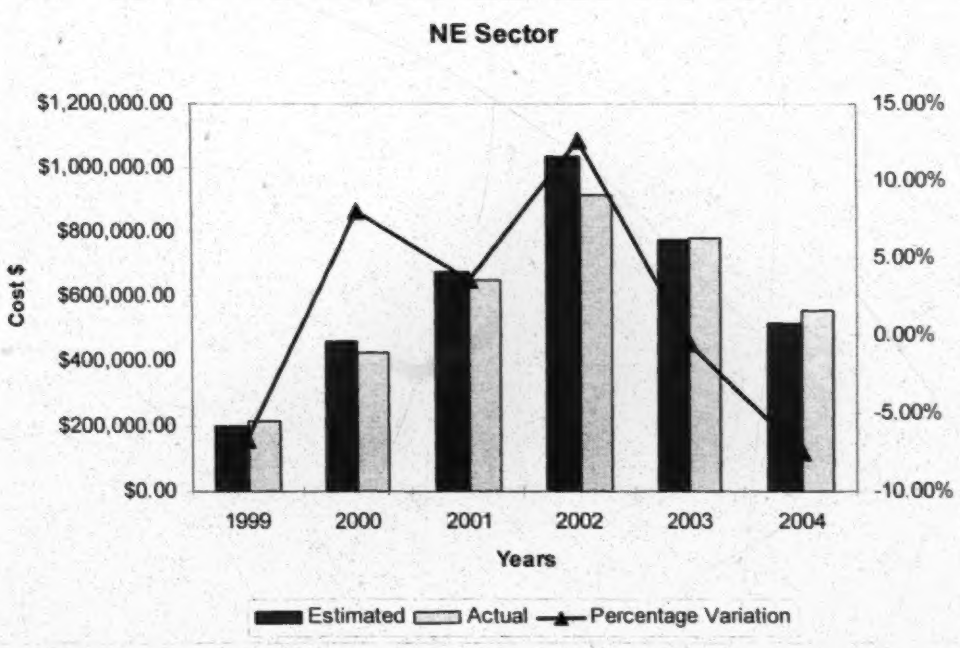



SW cost variation 2002
$25,000.00
$20,000.00
$15,000.00
$10,000.00
$5,000.00
$0.00
Cost
05/11/2001
13/02/2002
24/05/2002
01/09/2002
10/12/2002
20/03/2003
Date
Estimate
Actual


SW sector 2001
$60,000.00
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$0.00
Cost
01/10/2000
09/01/2001
19/04/2001
28/07/2001
05/11/2001
13/02/2002
Date
Estimate
Actual

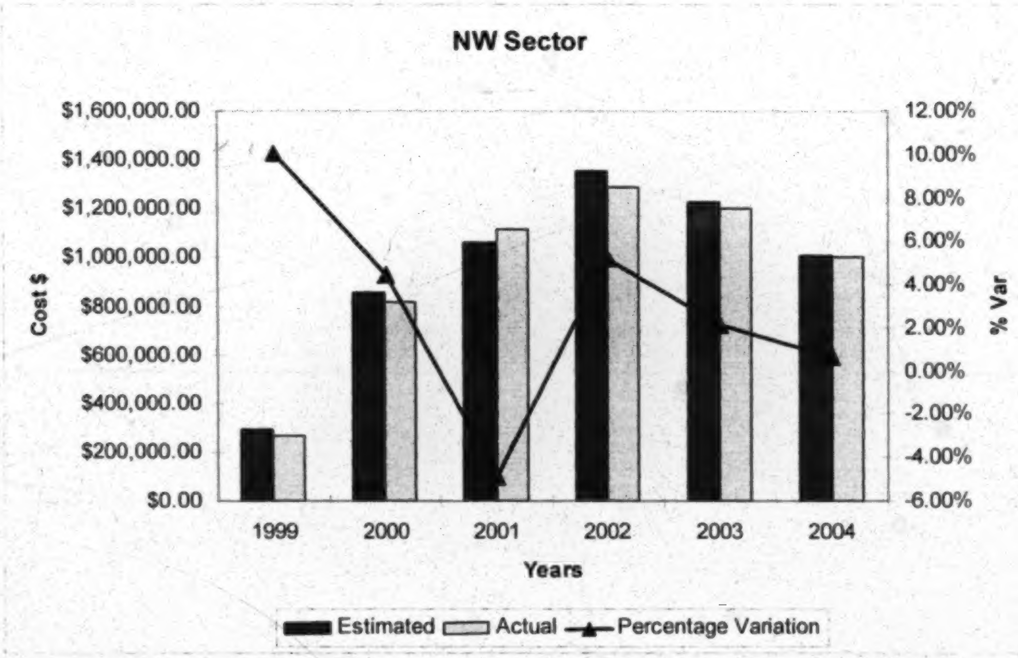

SW sector cost variation 2000
$70,000.00
$60,000.00
$50,000.00
$40,000.00
$30,000.00
$20,000.00
$10,000.00
$0.00
Cost
15/03/2000
04/05/2000
23/06/2000
12/08/2000
01/10/2000
20/11/2000
09/01/2001
Date
Estimate
Actual


SW sector cost variation1999
$7,400.00
$7,300.00
$7,200.00
$7,100.00
$7,000.00
$6,900.00
$6,800.00
Cost
Estimate
Actual
23/11/1999
Date
☐ Estimate
■ Actual



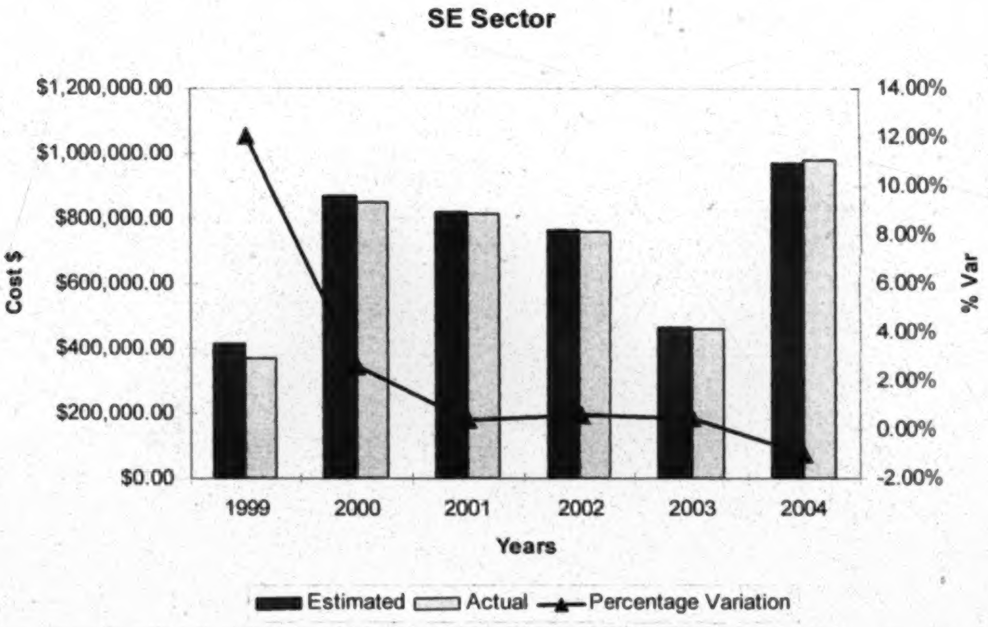
Year	NE*			
	Estimated	**Actual**	**Difference**	**Variation**
1999	$957,630.74	$1,025,733.49	($68,102.75)	-6.64%
2000	$2,179,115.72	$2,017,596.93	$161,518.80	8.01%
2001	$3,187,850.36	$3,077,546.06	$110,304.30	3.58%
2002	$4,897,527.62	$4,347,748.98	$549,778.64	12.65%
2003	$3,682,117.82	$3,701,788.84	($19,671.02)	-0.53%
2004	$2,431,169.66	$2,627,485.66	($196,316.00)	-7.47%

(* Note: Values within the table have been changed to protect the confidentiality of the Client)



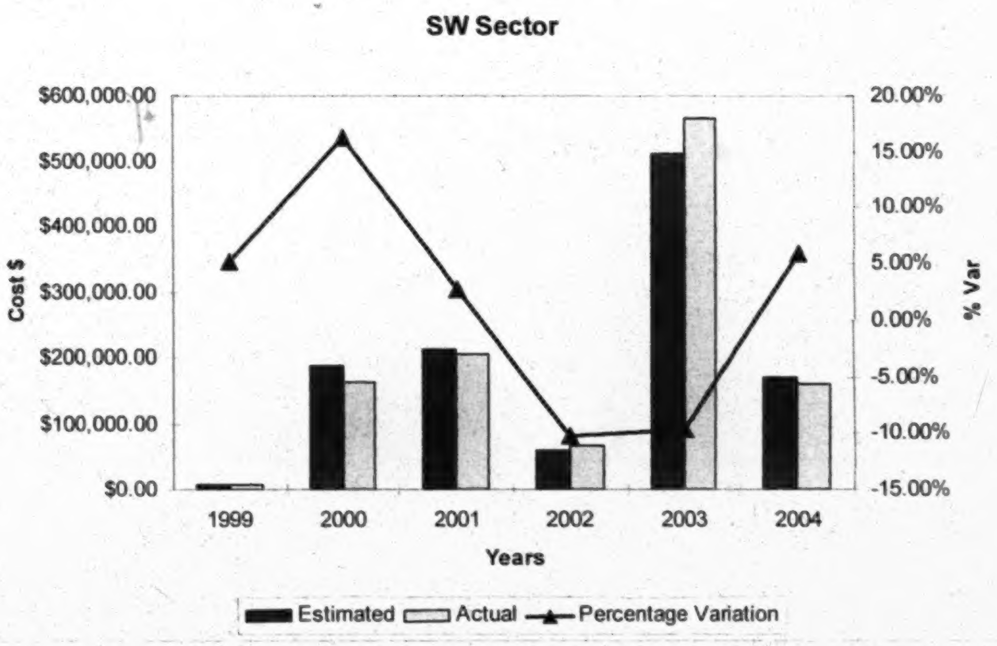
Year	NW*			
	Estimated	Actual	Difference	Variation
1999	$1,370,338.07	$1,244,785.43	$125,552.64	10.09%
2000	$4,052,755.02	$3,881,204.18	$171,550.84	4.42%
2001	$5,029,919.05	$5,288,129.57	($258,210.52)	-4.88%
2002	$6,415,395.59	$6,097,433.00	$317,962.59	5.21%
2003	$5,809,639.34	$5,688,498.03	$121,141.31	2.13%
2004	$4,797,008.49	$4,765,131.67	$31,876.82	0.67%

(* Note: Values within the table have been changed to protect the confidentiality of the Client)



Year	SE*			
	Estimated	Actual	Difference	Variation
1999	$1,969,013.40	$1,756,814.17	$212,199.23	12.08%
2000	$4,123,013.76	$4,018,454.21	$104,559.55	2.60%
2001	$3,883,817.69	$3,869,161.13	$14,656.56	0.38%
2002	$3,629,028.30	$3,608,292.82	$20,735.49	0.57%
2003	$2,200,264.49	$2,189,425.29	$10,839.19	0.50%
2004	$4,599,913.26	$4,647,005.61	($47,092.35)	-1.01%

(* Note: Values within the table have been changed to protect the confidentiality of the Client)


SW Sector

Year	SW*			
	Estimated	Actual	Difference	Variation
1999	$34,882.28	$33,182.70	$1,699.58	5.12%
2000	$893,658.29	$768,453.07	$125,205.22	16.29%
2001	$1,002,337.20	$975,601.35	$26,735.85	2.74%
2002	$282,092.28	$314,486.53	($32,394.26)	-10.30%
2003	$2,419,829.37	$2,678,185.27	($258,355.90)	-9.65%
2004	$805,434.45	$760,715.77	$44,718.67	5.88%

(* Note: Values within the table have been changed to protect the confidentiality of the Client)